

Gestörter Unterricht

Wie Lehrpersonen, Schülerinnen und Schüler Unterrichtsstörungen
wahrnehmen, deuten, bewerten und beschreiben

Abhandlung (kumulative Dissertation) zur Erlangung der Doktorwürde
der Philosophischen Fakultät der Universität Zürich
vorgelegt von Boris Eckstein

Eingereicht im Herbstsemester 2018 zuhanden der Promotionskommission

Prof. em. Dr. Kurt Reusser (hauptverantwortlicher Betreuer)

Prof. Dr. Elisabeth Moser Opitz (Koreferentin)

Prof. Dr. Alexander Wettstein (Koreferent)

Zürich, 2018

Zusammenfassung

Unterrichtsstörungen umfassen ein Element der Produktion (stören) sowie ein Element der Rezeption (gestört sein). Dieser Problemzusammenhang wurde im Rahmen der vorliegenden, kumulativen Dissertation in vier Originalarbeiten auf der Grundlage von drei Studien untersucht. In Beitrag I wurde ermittelt, dass Lehrpersonen die Realisierbarkeit der integrativen Schulungsform zunehmend skeptisch einschätzen, je mehr Schüler/innen mit Verhaltensauffälligkeiten in ihrer Klasse sind. Dieser Effekt lässt sich theoretisch damit erklären, dass Lehrpersonen bestimmte Verhaltensweisen dieser Schüler/innen im Unterricht als störend empfinden. Beitrag II ergänzt diesen Befund mit Resultaten einer zweiten Studie: Schülerverhalten, welches von unterrichtlichen Normen abweicht (z.B. Undiszipliniertheit), wird nicht von allen beteiligten Akteuren als gleichermassen störend empfunden. Diesen Befund vertiefen die Beiträge III und IV basierend auf einer dritten Studie weiter. Deskriptive Resultate legen nahe, dass sich Unterrichtsstörungen im Durchschnitt selten ereignen. Darüber hinaus wurden mittels eines Strukturgleichungsmodells des Typs CT-C(M-1) Rater-Effekte ermittelt, welche die Wahrnehmung, Deutung und Bewertung gestörten Unterrichts zu einem substanziellen Anteil erklären. Lehrpersonen registrieren z.B. mehr Undiszipliniertheiten als Schüler/innen. Im Manteltext der Dissertation werden die vier Beiträge zunächst in einem interaktionistischen Theorierahmen situiert und hernach in einer abschliessenden Gesamtbetrachtung hinsichtlich forschungsmethodischer und praxisrelevanter Implikationen diskutiert.

Schlagworte: Unterrichtsstörungen; normabweichendes Schülerverhalten; subjektives Störungsempfinden; interaktionistische Theorie, CT-C(M-1) Modell

Abstract

Classroom disturbances consist of an objective core that can be differently perceived, interpreted, and judged from the perspectives of the persons involved. The present cumulative dissertation focuses on this interactionist problem in four articles which base on three studies. Article I has shown that teachers are increasingly sceptical towards inclusion when the number of students with behavioural problems in their class increases. This effect can be explained theoretically: the teachers perceive certain behaviours of those students as disturbing. Article II complements this finding with results of a second study: Not everyone in class gets equally distracted or annoyed by the same occurrences of deviant student behaviour (e.g. indiscipline). Articles III and IV base on a third study which deepens the above mentioned findings. Descriptive Results suggest that classroom disturbances on average occur rather rarely. Beyond that, a CT-C(M-1) structural equation model detected rater effects that influence the perception of classroom disturbances substantially. Teachers, for instance, notice more occurrences of indiscipline than students. To begin with, the present synopsis situates the four articles in an interactionist theoretical framework. In a conclusive chapter, the most important findings are summarised and discussed in regard to scientific and practical implications.

Keywords: Classroom disturbances; deviant behaviour; subjective perception; interactionist theory; quantitative analysis

Inhaltsverzeichnis

Zusammenfassung	3
Abstract.....	4
Inhaltsverzeichnis	5
Widmung und Dank.....	8
1. Einleitung zur kumulativ angelegten Dissertation	10
1.1. Problemstellung, Erkenntnisinteresse und Ziele der Arbeit	11
1.2. Entstehungskontext, Datengrundlage und Aufbau der Arbeit	14
1.3. Vorbemerkungen zu formalen Aspekten der Arbeit	17
2. Übersicht über die vier Originalarbeiten.....	18
2.1. Beitrag I: Schüler/innen mit Verhaltensauffälligkeiten in der Regelschule	18
2.1.1. Bildungspolitischer Hintergrund: Die integrative Volksschule	19
2.1.2. Das Heterogenitäts-Projekt.....	20
2.1.3. Bedingungen optimistischer Überzeugungen von Lehrpersonen	22
2.1.4. Skepsis gegenüber Schülerinnen und Schülern mit Verhaltensauffälligkeiten	26
2.1.5. Schwierigkeiten bei der Ausmarchung des Problemausmasses.....	28
2.1.6. Grenzen der individualpsychologischen Suchlinie	31
2.2. Beitrag II: Gestörter Unterricht – Entflechtung zentraler Merkmalsdimensionen	33
2.2.1. Das SUGUS-Projekt	34
2.2.2. Schülerverhalten im Spiegel unterrichtlicher Normen	36
2.2.3. Unterrichtliche Devianz: Ätiologie, Prävention, Intervention.....	41
2.2.4. Unterrichtliche Devianz als Folge von Etikettierung?	48
2.2.5. Subjektives Störungsempfinden: Wahrnehmungspsychologische Ansätze	50
2.2.6. Theorieintegration: Interaktionistisches Modell gestörten Unterrichts.....	59

2.2.7.	Konzeptualisierung des Forschungsdesigns der quantitativen SUGUS-Studie.....	61
2.2.8.	Entwicklung und Pretesting der Instrumente.....	63
2.2.9.	Faktorenanalytische Untersuchung der im Pretest eingesetzten Instrumente.....	65
2.2.10.	Deskriptive Befunde der Vorstudie.....	69
2.2.11.	Zentrale Schlussfolgerungen und offen gebliebene Fragen nach der Vorstudie.....	73
2.3.	Beitrag III: Deskriptiver Perspektivenvergleich	76
2.3.1.	Vorbemerkungen zur SUGUS-Hauptstudie	77
2.3.2.	Sampling bei der Hauptstudie.....	78
2.3.3.	Weiterentwicklung des Instrumentariums nach der Vorstudie.....	80
2.3.4.	Datencleaning und vorbereitende deskriptive Analysen.....	83
2.3.5.	Perspektivenvergleich innerhalb der an der Studie teilnehmenden Schulklassen.....	86
2.3.6.	Grenzen des deskriptiven Perspektivenvergleichs.....	95
2.4.	Beitrag IV: Konsistenz und Spezifität der verschiedenen Perspektiven	97
2.4.1.	Skalierung der Konstrukte Devianz und Störungsempfinden in je zwei Dimensionen.....	98
2.4.2.	Varianzanalytischer Perspektivenvergleich anhand der vier Konstrukte	100
2.4.3.	Perspektivenvergleich anhand des CT-C(M-1) Ansatzes.....	106
3.	Zusammenfassung und Bilanz	117
3.1.	Haupterträge der einzelnen Originalarbeiten.....	118
3.1.1.	Beitrag I: Schüler/innen mit Verhaltensauffälligkeiten in der Regelschule	118
3.1.2.	Beitrag II: Gestörter Unterricht – Entflechtung zentraler Merkmalsdimensionen.....	119
3.1.3.	Beitrag III: Deskriptiver Perspektivenvergleich	120
3.1.4.	Beitrag IV: Konsistenz und Spezifität der verschiedenen Perspektiven	121
3.2.	Übergeordnete Erträge der vier Originalarbeiten.....	124
3.2.1.	Methodologische Implikationen für die Erforschung von Unterrichtsstörungen.....	124
3.2.2.	Implikationen für die Praxis	128

3.2.3. Ethische Implikationen in Bezug auf die Diagnostik von "Problemschülern"	129
3.3. Limitationen der Arbeit	130
3.4. Ausblick auf weitere Analysen im SUGUS-Projekt	134
4. Literaturverzeichnis	141
5. Tabellen- und Abbildungsverzeichnis	155
6. Anhang.....	156
6.1. Originalarbeiten	156
6.1.1. Umsetzung der integrativen Volksschule – Was Lehrpersonen optimistisch macht	156
6.1.2. Unterrichtliche Devianz und subjektives Störungsempfinden.....	187
6.1.3. Unterrichtsstörungen: eine Frage der Perspektive?.....	208
6.1.4. Production and Perception of Classroom Disturbances	226
6.2. Curriculum Vitae.....	260
6.3. Selbstständigkeitserklärung	261

Widmung und Dank

Neben meinem Studium habe ich als Sozialpädagoge in einem sonderpädagogischen Schulinternat gearbeitet. Dabei habe ich erfahren, wie unterschiedlich wir Pädagoginnen und Pädagogen auf schwieriges Verhalten der Klientinnen und Klienten reagierten. Ich persönlich war im Allgemeinen eher empfindlich und habe herausfordernde Situationen oftmals als belastender erlebt als manche Kolleginnen und Kollegen. Zum *Klienten O.* aber hatte ich eine besondere Beziehung und konnte ihm viel Verständnis entgegenbringen. Er tat mir vor allem leid wegen seiner schwierigen Kindheit und ich hatte oftmals den Eindruck, er wolle gar nicht streiten und stören, könne aber nicht anders. So fiel mir die erzieherische Arbeit mit O. manchmal etwas leichter als anderen Mitarbeitenden. Meine Dissertation hat nur zwar am Rande mit dieser Problematik zu tun, ist aber thematisch mit ihr verwandt, sodass ich sie O. und allen anderen Kindern und Jugendlichen widme, die eigentlich nicht stören wollen, manchmal aber nicht anders können.

Gegenstand meiner Dissertation sind *Unterrichtsstörungen* an der Regelschule, welche meistens weniger gravierend sind als Störungen, welche sich im oben erwähnten sonderpädagogischen Heimkontext manchmal ereignen. Allerdings wird bisweilen unterstellt, die Problematik alltäglicher Unterrichtsstörungen habe sich in der jüngeren Vergangenheit verschärft. Vor diesem Hintergrund interessieren mich die Auftretenshäufigkeit und der Schweregrad unterrichtlicher Störungen sowie deren differenzielle Wahrnehmung, Deutung und Bewertung durch Lehrpersonen und Schüler/innen. Dieses Forschungsinteresse entwickelte sich in der Zusammenarbeit mit von mir sehr geschätzten Wissenschaftlerinnen und Wissenschaftlern, denen ich an dieser Stelle danken möchte.

Meinem Doktorvater, Herrn Prof. em. Dr. Kurt *Reusser*, danke ich für seine substanzielle Förderung meiner akademischen Entwicklung sowie für seine vertrauensvolle und konstruktive Betreuung meines Promotionsvorhabens. Ich danke auch der Koreferentin meiner Dissertation, Frau Prof. Dr. Elisabeth *Moser Opitz*, die stets niederschwellig zu einem Austausch über fachliche und überfachliche Themen des forschungspraktischen Alltags bereit war. Herrn Prof. Dr. Alexander *Wettstein*, dem zweiten Koreferenten, danke ich für anregende Fachgespräche, welche die theoretischen und methodischen Grundlagen der SUGUS Studie geprägt haben.

Bei Herrn Dr. Urs *Grob* bedanke ich mich für seine freundschaftliche Unterstützung und Beratung bei der Planung der Datenerhebung und Auswertungsstrategie sowie bei der Durchführung der statistischen Analysen und bei der Interpretation der Resultate – die zahlreichen Gespräche habe ich als grosse Hilfe und Bereicherung erlebt. Frau Dr. Rita *Stebler* und Frau lic. phil. Debbie *Mandel* danke ich für die angenehme Zusammenarbeit im Heterogenitäts-Projekt. Bei meinen Kolleginnen Dr. Katriina *Vasarik Staub*, MSc Anna *Hofstetter* und MA Claudia *Marusic-Würscher* bedanke ich mich für den bereichernden Austausch über unsere Forschung im SUGUS-Projekt. Herrn MA Simon *Luger* und zahlreichen weiteren *studentischen Hilfskräften* danke ich für ihre Mitarbeit in der quantitativen SUGUS-Studie. Bei Herrn Prof. Dr. Christoph *Müller* bedanke ich mich für seine Beratung bei der Konzeptualisierung des SUGUS-Projekts. Den Herren Prof. Dr. Martin *Venetz* (†) und Prof. Dr. Fridtjof *Nußbeck* danke ich für ihre Unterstützung bei der CT-C(M-1) Analyse.

Ein spezieller Dank gilt meinen zwei allerliebsten Töchtern *Smilla und Maja*, welche das zuweilen knappe Zeitbudget ihres promovierenden Papas mit Nachsicht geteilt haben. Und natürlich bedanke ich mich bei meiner wundervollen Frau *Sabine*, die mich in dieser intensiven Phase fortwährend liebevoll gestärkt hat.

1. Einleitung zur kumulativ angelegten Dissertation

Diese Klasse ist untereinander im Umgang sehr liebenswürdig und das Unterrichten mit ihnen sehr angenehm. (Klasse 019, Lehrperson)

Aber wenn unsere Lehrerin raus geht sind die einten sehr laut und schreien schon fast dann bekomme ich immer Kopfweh. Das nervt dann sehr fest.
(Klasse 019, Schüler/in 04)

Unsere Lehrerin versteht sehr viel etwas falsch z.b Ich und Klassenkamerad spielen und schlagen uns aus Spass, mir und ihm macht es nichts aus doch sie macht ein riesen Drama! (Klasse 019, Schüler/in 01)

Ich hasse es wen wir z.B. in der Mathe leise was lösen müssen und die anderen schwatzen die ganze zeit. (Klasse 019, Schüler/in 17)

Mich stört starkes Atmen. (Klasse 019, Schüler/in 18)

Diese fünf Zitate stammen von einer Lehrperson und von vier ihrer Schüler/innen, welche im Rahmen der *"Studie zur Untersuchung gestörten Unterrichts (SUGUS)"* einen Fragebogen beantworteten (Eckstein, Luger, Grob & Reusser, 2018). Es handelt sich um handschriftliche Kommentare, mit denen die Befragten ihre zuvor standardisiert erfassten Angaben im freien Format ergänzten. Die Kommentare wurden bislang nicht systematisch ausgewertet, daher dienen die Zitate in erster Linie als Anekdote. Ich möchte damit folgendes Kernargument meiner Dissertation veranschaulichen: *Es gibt verschiedene Formen von Unterrichtsstörungen, welche in der Selbst- und Fremdwahrnehmung von Lehrpersonen, Schülerinnen und Schülern perspektivisch unterschiedlich gedeutet und bewertet werden können.* Dieses Argument werde ich in der vorliegenden Arbeit begründen und vertiefen.

1.1. Problemstellung, Erkenntnisinteresse und Ziele der Arbeit

Unterrichtsstörungen zeichnen sich dadurch aus, dass die Lehr-Lernprozesse ins Stocken geraten, temporär unterbrochen werden oder komplett abbrechen (Hofstetter & Reusser, 2016; Winkel, 2011). Damit einher gehen oftmals auch *negative Auswirkungen* auf das emotional-affektive Empfinden der Beteiligten. Aus diesen Gründen sind Unterrichtsstörungen eine potenzielle Belastungsquelle für Lehrpersonen (Kokkinos, 2007; Krause, Dorsemagen & Alexander, 2011) sowie für Schüler/innen (Grosse Siestrup, 2008; Infantino & Little, 2005).

Als proximale Ursachen gestörten Unterrichts gelten unangemessene Verhaltensweisen von Schülerinnen, Schülern oder Lehrpersonen sowie externe Quellen (Menzel, 2009; Montuoro & Lewis, 2015; Wicki & Kappeler, 2007). Um eine Störung handelt es sich aber erst dann, wenn mindestens eine am Unterricht beteiligte Person das Ereignis als störend wahrnimmt, deutet und bewertet (Eckstein, Grob & Reusser, 2016). Gemäss diesem interaktionistischen Problemverständnis (Schweer & Thies, 2000) umfassen Unterrichtsstörungen zwei Hauptelemente: (1) Produktion – stören; (2) Rezeption – gestört sein.

In der Literatur wurden und werden Unterrichtsstörungen unter dem Aspekt der Produktion sehr oft als primär schülerseitiges Problem konzeptualisiert. Weil sich manche Schüler/innen häufiger unangemessen verhalten als andere, untersuchten viele Studien ontogenetisch bedingte Risikofaktoren störenden Verhaltens, z.B. Impulsivität (Carroll, Houghton, Taylor, West & List-Kerz, 2006). Gemäss dieser personenzentrierten (individualpsychologischen) Sichtweise erhalten Schüler/innen mit Verhaltensauffälligkeiten eine besondere Aufmerksamkeit. Auch vor dem Hintergrund der bildungspolitischen Entwicklung in Richtung einer integrativen Volksschule wird dieser Aspekt des Problemzusammenhangs in akzentuierter Weise diskutiert (Bernhardt, 2007; Schwab, 2015).

Eine andere Theorie- und Forschungstradition rückt demgegenüber das differenzielle Störungsempfinden von Lehrpersonen, Schülerinnen und Schülern in den Vordergrund. Untersucht werden bei diesem Zugang beispielsweise die Wahrnehmung, Deutung und Bewertung von Unterrichtsstörungen aus unterschiedlichen Perspektiven (Wettstein, Scherzinger & Ramseier, 2018).

In der vorliegenden, kumulativ angelegten Dissertation unternehme ich den Versuch, die beiden Zugänge zu integrieren und damit zu einer stärker ganzheitlichen Problemsicht beizutragen. Der verfolgte integrative Ansatz besteht dabei aus den folgenden drei *theoretischen Hauptargumenten*: (1) Bestimmtes Schülerverhalten stellt in bestimmten Situationen eine Abweichung von unterrichtlichen Normen dar – diese unterrichtliche Devianz kann als objektiver Kern von Unterrichtsstörungen aufgefasst werden. (2) Lehrpersonen und Schüler/innen nehmen deviantes Schülerverhalten in ihren jeweiligen subjektiven Perspektiven als mehr oder weniger stark störend wahr. (3) Sowohl die Auftretenshäufigkeit devianten Verhaltens als auch die Intensität und die Qualität des subjektiven Störungsempfindens werden von verschiedenen Bedingungen beeinflusst.

Ausgehend von dieser in der Arbeit auch als ‘ganzheitlich’ bezeichneten theoretischen Konzeptualisierung gestörten Unterrichts gehe ich in der Dissertation auch auf *methodologische Grundlagen* ein, die es ermöglicht haben, Gemeinsamkeiten und Unterschiede zwischen den Perspektiven von Lehrpersonen, Schülerinnen und Schülern zu untersuchen. Dabei werde ich u.a. ein neuartiges statistisches Verfahren zur Analyse komplexer Multitrait-Multimethod (MTMM) Daten vorstellen: ein Mehrebenen-Strukturgleichungsmodell des Typs CT-C(M-1) (Eid et al., 2008). Dieses Verfahren hat es mir erlaubt, die Schnittmenge der Perspektiven von

Lehrpersonen, Schülerinnen und Schülern sowie subjektive Abweichungen von dieser intersubjektiv geteilten Sichtweise zu ermitteln.

Das bedeutet, Unterrichtsstörungen werden in der vorliegenden Dissertation als interaktional ko-konstruierte Phänomene aufgefasst und mit entsprechenden methodischen Zugängen erforscht. Die Ausführungen in dieser Mantelschrift beruhen auf vier aufeinander aufbauenden (im Anhang dokumentierten) Originalarbeiten, in denen ich mehrere *Teilaspekte des Problemzusammenhangs* "gestörter Unterricht" untersucht habe. Die konkreten Fragestellungen werden in den nachfolgenden Kapiteln 2.1 bis 2.4 bei der Darstellung der einzelnen Beiträge genannt. Zusammenfassend lässt sich festhalten, dass die unternommenen Analysen folgende *drei Hauptziele* verfolgten:

1. Analyse und kritische Diskussion des Einflusses schülerseitiger Verhaltensauffälligkeit auf integrationsbezogene Überzeugungen von Lehrpersonen (Beitrag I)
2. Ganzheitliche (Produktions- und Rezeptionsmerkmale integrierende) Herangehensweise an die theoretische Konzeptualisierung und empirische Untersuchung gestörten Unterrichts (Beitrag II)
3. Erforschung von Gemeinsamkeiten und Unterschieden in der Wahrnehmung, Deutung und Bewertung von Unterrichtsstörungen durch Lehrpersonen, Schülerinnen und Schüler
 - a. Deskriptiver Perspektivenvergleich (Beitrag III)
 - b. Konsistenz und Spezifität der verschiedenen Perspektiven (Beitrag IV)

Ziel des vorliegenden Manteltexts ist es, die in den vier Beiträgen gewonnenen Erkenntnisse – breiter als dies das Format der Originalarbeiten erlaubte – theoretisch einzuordnen und zu einem *Gesamtbild* zu integrieren sowie mit weiterführenden Überlegungen zu ergänzen und zu diskutieren.

1.2. Entstehungskontext, Datengrundlage und Aufbau der Arbeit

Bei den vier Originalarbeiten der Dissertation handelt sich um drei Beiträge für wissenschaftliche Zeitschriften sowie um einen Beitrag für einen Tagungsband, die alle einen Peer-Review-Prozess durchliefen. Zum gegenwärtigen Zeitpunkt sind die Beiträge I bis III publiziert; Beitrag IV ist eingereicht. Die vier Originalarbeiten entstanden im Rahmen zweier Forschungsprojekte auf der Grundlage von *drei Studien*.

Beitrag I (Eckstein, Reusser, Stebler & Mandel, 2013) entstand im Rahmen des Projekts *"Erfolgreicher Unterricht in heterogenen Lerngruppen auf der Volksschulstufe des Kantons Zürich"* (kurz: Heterogenitäts-Projekt; Laufzeit: 2010-2013). Hierbei handelt es sich um eine Auftragsstudie, welche von Prof. em. Dr. Kurt Reusser geleitet wurde. Die Studie umfasste mehrere Module, darunter qualitative Leitfaden-Interviews und eine standardisierte Online-Befragung. Meine Funktion im Heterogenitäts-Projekt bestand in der Mitarbeit bei der Vorbereitung und Durchführung der Online-Umfrage sowie bei der Verfassung des wissenschaftlichen Berichts (Reusser, Stebler, Mandel & Eckstein, 2013).

Beitrag I (Eckstein et al., 2013) beruht auf einer vertiefenden Analyse der aus der Online-Befragung gewonnenen Daten zu integrationsbezogenen Überzeugungen von Lehrpersonen. Bei diesem Artikel habe ich als Erstautor die Analysen geplant und durchgeführt sowie den Manuskriptentwurf verfasst und nach internem Feedback der Ko-Autorinnen und des Ko-Autors sowie gemäss den anonymen Peer-Reviews überarbeitet.

Ein Ergebnis der in Beitrag I unternommenen Analyse hat neue Fragen aufgeworfen, welche zur Ausrichtung eines zweiten Projekts beigetragen haben. Damit ist die *"Studie zur*

Untersuchung gestörten Unterrichts" (kurz: SUGUS) angesprochen. Die SUGUS-Studie ist ein vom Schweizerischen Nationalfonds (SNF) gefördertes Forschungsprojekt (Projektnummer: 100019_152722; Laufzeit: 2014 – 2019), welches von Prof. em. Dr. Kurt Reusser geleitet wird. Das Projekt besteht aus einer qualitativen und aus einer quantitativen Teilstudie. In der Konzeptphase des Projekts bestand meine Funktion darin, unter der Anleitung von Kurt Reusser einen Antragsentwurf zu verfassen. Nach Bewilligung des Projekts habe ich als vom SNF geförderter Doktorand die operative Durchführung der quantitativen SUGUS-Teilstudie geleitet.

Beitrag II erfolgte im Rahmen der *SUGUS-Vorstudie*, in der das Studiendesign und das Instrumentarium der späteren Hauptstudie einem Pretest unterzogen wurden (Eckstein, Grob, et al., 2016). Wie bei Beitrag I habe ich auch bei diesem Artikel als Erstautor die Analysen geplant und durchgeführt sowie den Manuskriptentwurf verfasst und nach internem Feedback der Ko-Autoren sowie gemäss den anonymen Peer-Reviews überarbeitet.

Beitrag III (Eckstein, 2018) und **Beitrag IV** (Eckstein, under review) basieren auf der *SUGUS-Hauptstudie*, an der sich 85 Schulklassen beteiligten. Bei diesen zwei Beiträgen habe ich als Alleinautor die Analysen geplant und durchgeführt sowie den Manuskriptentwurf verfasst und gemäss den anonymen Peer-Reviews überarbeitet.

Im folgenden Kapitel 2 werde ich problementwickelnd darlegen, wie die Beiträge zusammenhängen und aufeinander aufbauen. Dazu werde ich die Originalarbeiten in einer *chronologisch aufgebauten Übersicht* jeweils in einem eigenen Abschnitt zusammenfassen und mit weiterführenden Informationen anreichern.

Die theoretische Auseinandersetzung mit dem Themenkomplex erfolgt schwerpunktmässig in den Kapiteln 2.1 und 2.2. Hier wird auch die *Schnittstelle zwischen den zwei Forschungsprojekten* bzw. deren innerer Zusammenhang dargelegt. Um die beiden Projekte mit einer kohärenten Argumentationslinie zu verknüpfen, greife ich in Kapitel 2.1 ein spezifisches Resultat von Beitrag I heraus und vertiefe dieses mit weiterführenden, literaturbasierten Argumenten: Es geht dabei um eine im (fach-)öffentlichen Diskurs – auch unter Lehrpersonen – weit verbreitete Skepsis gegenüber der Integration von Schülerinnen und Schülern mit Verhaltensauffälligkeiten sowie um damit einhergehende Unklarheiten theoretischer und forschungsmethodischer Art. Diese Ausführungen werden im anschliessenden Kapitel 2.2, in dem die theoretischen und methodischen Grundlagen des SUGUS-Projekts sowie zentrale Befunde der SUGUS-Vorstudie präsentiert werden, aufgegriffen und fortgeführt.

Die Kapitel 2.3 und 2.4 sind vorrangig den empirischen Resultaten der SUGUS Hauptstudie gewidmet, welche ich dort mit bislang unveröffentlichten Ergebnissen vorbereitender und weiterführender Analysen ergänze und vertiefe.

In Kapitel 3 folgt eine abschliessende *Gesamtdiskussion*. Darin stelle ich die Hauptergebnisse der Analysen nochmals in verdichteter Form dar und bilanziere übergeordnete Erträge der Dissertation als Ganzes. Daraufhin thematisiere ich Limitationen der vier Beiträge und nehme einen Ausblick auf mögliche weiterführende Analysen vor. Die vier Originalarbeiten finden sich im Anhang ab Seite 156.

1.3. Vorbemerkungen zu formalen Aspekten der Arbeit

Orientierung an APA- und DGPs-Normen. Bei der formalen Gestaltung der vorliegenden Arbeit habe ich mich an den Richtlinien der American Psychological Association (APA) orientiert (American Psychological Association, 2010). Aufgrund der deutschen Sprache der Arbeit habe ich Quellennachweise gemäss den Zitationsrichtlinien der Deutschen Gesellschaft für Psychologie (DGPs) erbracht (Deutsche Gesellschaft für Psychologie, 2007), welche ihrerseits auf den APA-Normen basieren.

Geschlechtsneutrale Schreibweise. In sprachlicher Hinsicht habe ich mich bemüht, vorrangig geschlechtsneutrale Bezeichnungen zu verwenden, z.B. Lehrpersonen. Die "Schülerinnen und Schüler" werden zugunsten der sprachlichen Verdichtung auch als "Schüler/innen" bezeichnet, ausser die Grammatik verlangt nach einer damit unverträglichen Flexion. Bei Komposita, z.B. Schülerverhalten, wird auf eine die Lesbarkeit erschwerende Neutralisierung des grammatischen Genus verzichtet. Ebenfalls unverändert bleiben Anglizismen wie "Rater".

Originale Schreibweise von Befragten. In der Arbeit zitiere ich mehrfach Kommentare von Lehrpersonen, Schülerinnen und Schülern, welche sie im Fragebogen notiert hatten (z.B. am Anfang von diesem Kap 1). Zwecks Authentizität habe ich die Kommentare in originaler Schreibweise unverändert übernommen, d.h. Rechtschreibfehler und dgl. wurden nicht korrigiert.

2. Übersicht über die vier Originalarbeiten

2.1. Beitrag I: Schüler/innen mit Verhaltensauffälligkeiten in der Regelschule

Abstract aus der Originalarbeit

Seit der Annahme des neuen Zürcher Volksschulgesetzes im Jahr 2005 ist das kantonale Bildungssystem dazu verpflichtet, möglichst alle Lernenden mit besonderen pädagogischen Bedürfnissen integrativ in Regelklassen zu unterrichten. Die Lehrpersonen sind aufgefordert, in heterogenen Lerngruppen Unterricht zu gestalten, der Schüler/innen mit unterschiedlichsten Voraussetzungen und Bedürfnissen optimal fördert. Der Artikel geht der Frage nach, unter welchen Voraussetzungen Lehrpersonen optimistisch sind, dass die Umsetzung der integrativen Volksschule gelingt. Die im Rahmen des Heterogenitäts-Projekts (Universität Zürich, Institut für Erziehungswissenschaft) via eine standardisierte Online-Befragung bei 385 Lehrpersonen aus 19 Schulen im Kanton Zürich erhobenen Daten werden genutzt, um in einem Strukturgleichungsmodell den Einfluss von vier Prädiktoren auf das latente Konstrukt ‚Integrationsoptimismus‘ zu analysieren. Die Ergebnisse zeigen, dass sich der Integrationsoptimismus der Lehrpersonen vor allem durch eine Erweiterte Didaktische Unterrichtspraxis sowie die Quote verhaltensauffälliger Schüler/innen pro Klasse vorhersagen lässt. In der Diskussion werden das Auswertungsverfahren und die praktische Relevanz der Befunde reflektiert. (Eckstein et al., 2013, S. 91)¹

Um diesen ersten Beitrag meiner kumulativen Dissertation in einem grösseren Zusammenhang darzustellen, werde ich in einem ersten Schritt den Entstehungskontext beschreiben (Kap. 2.1.1, Kap. 2.1.2). Sodann präsentiere ich in Kapitel 2.1.3 die Ergebnisse der Analyse. In den nachfolgenden Kapiteln (2.1.4 – 2.1.6) diskutiere ich einen Hauptbefund unter Bezugnahme theoretischer und forschungsmethodischer Argumente.

¹ Die vollständigen bibliographischen Angaben der Originalarbeit lauten: Eckstein, B., Reusser, K., Stebler, R. & Mandel, D. (2013). Umsetzung der integrativen Volksschule – Was Lehrpersonen optimistisch macht: Eine Analyse der Überzeugungen von Klassenlehrpersonen im Kanton Zürich. *Schweizerische Zeitschrift für Bildungswissenschaften*, 35(1), 91-112.

2.1.1. Bildungspolitischer Hintergrund: Die integrative Volksschule

Im deutschsprachigen Raum sind viele Grundstrukturen des Schulwesens historisch bedingt mehrgliedrig, segregativ organisiert (Allemann-Ghionda, 2008; Jenzer, 1998; Prengel, 2010; Wolfisberg, 2008). Erst einem jüngeren, internationalen Trend (UNESCO, 1994) folgend haben die gesetzgebenden Instanzen in Deutschland, Österreich und in der Schweiz beschlossen, Kinder und Jugendliche mit besonderen pädagogischen Bedürfnissen vermehrt *integrativ in Regelschulen* statt wie bisher vornehmlich separativ in Förderschulen oder Sonderklassen zu schulen. Damit wurde auf der Systemebene des Bildungswesens ein Reformprogramm mit potenziellen Tiefenwirkungen angestoßen, wobei notwendige Entwicklungsschritte in verschiedenen Bereichen (z.B. Infrastruktur, Lehrmittel, Personal) noch nicht überall optimal realisiert worden sind (Katzenbach, 2017).

Die Umsetzung dieser Reform dürfte mancherorts auch deshalb zögerlich verlaufen, weil *Integration nicht nur unkritisch gesehen wird* (Tillmann, 2008; Trautmann & Wischer, 2011). Eine kontrovers diskutierte Frage betrifft etwa den anzustrebenden Inklusionsgrad (Poscher, Langer & Rux, 2008): Wie viele Schüler/innen mit welchen besonderen pädagogischen Bedürfnissen kann/soll/muss die Volksschule integrieren? Diese Frage spiegelt sich beispielsweise in einer Debatte über die korrekte Bezeichnung – Integration vs. Inklusion².

² Boban und Hinz (2009, 2012) favorisieren den in Deutschland mittlerweile gebräuchlicheren Begriff "Inklusion", weil dieser die umfassendste Form des Einschlusses bezeichne und zugleich positiv konnotiere. Dagegen wird in der Deutschschweiz oftmals der Begriff "Integration" verwendet (Bless, Sermier Dessemonet & Benoit, 2010; Kunz, Luder & Moretti, 2010; Liesen & Luder, 2011), womit aus hiesiger Sicht nicht etwas grundlegend Anderes gemeint ist. In der vorliegenden Dissertation kann keine detaillierte Begriffsklärung geleistet werden, weswegen ich die beiden Begriffe im Folgenden synonym verwenden werden.

2.1.2. Das Heterogenitäts-Projekt

Mit der Annahme des neuen Volksschulgesetzes wurde die integrative Schulung und damit der Auftrag, Schülerinnen und Schüler wenn immer möglich und vertretbar in Regelklassen zu unterrichten, auch im Kanton Zürich rechtlich verankert (Kantonsrat des Kantons Zürich, 2005). Vor diesem Hintergrund gab die kantonale Bildungsdirektion die Studie "Erfolgreicher Unterricht in heterogenen Lerngruppen auf der Volksschulstufe des Kantons Zürich" (kurz: Heterogenitäts-Projekt; Laufzeit: 2010-2013) in Auftrag (Reusser et al., 2013). Die Studie bestand aus qualitativen Erhebungen an sechs Fallschulen (darunter Leitfaden-Interviews mit Lehrpersonen) sowie aus einer standardisierten Online-Befragung von 385 Lehrpersonen aus 19 Schulen im Kanton Zürich (inklusive der sechs Fallschulen). In Form einer wissenschaftlichen Expertise zu theoretischen Grundlagen und zum Forschungsstand sowie einer empirischen Untersuchung sollten Befunde bezüglich Unterrichtsqualität in integrativ unterrichteten (heterogenen) Lerngruppen erarbeitet, nebst Qualitäten auch Defizite identifiziert und ggf. Massnahmen zur Unterstützung vorgeschlagen werden.

Ein wichtiges Ergebnis der Studie war, dass die Mehrheit der Lehrpersonen integrative Schulung möglichst aller Kinder *zwar für wünschbar – nicht aber für machbar* hält unter den gegebenen Voraussetzungen. Zwar erachten die Lehrpersonen die integrierten Schüler/innen grundsätzlich als *Bereicherung* für die Regelschule, doch erleben sie die Integration auch als *Belastung und Überforderung* (Reusser et al., 2013). Diese Ambivalenz lässt sich teilweise mit einem von den Lehrpersonen wahrgenommenen Ressourcenproblem erklären (Mandel, in Vorbereitung).

Von den Lehrpersonen wird erwartet, dass sie die heterogenen Voraussetzungen und Bedürfnisse ihrer Schüler/innen bei der Unterrichtsplanung proaktiv berücksichtigen und die Lehr-Lernprozesse möglichst adaptiv gestalten, etwa mittels binnendifferenzierender Massnahmen (Joller-Graf, 2010; Weinert, 1997). Eine solche Unterrichtsgestaltung stellt *hohe Anforderungen* an die Lehrpersonen (z.B. Diagnostik; Überblick über mehrere simultan stattfindende Lernaktivitäten) und bedarf einer zeitintensiven Vorbereitung (Trautmann & Wischer, 2008). Das war bereits vor der Integration so, jedoch ist der Anspruch durch diese noch deutlicher ins Bewusstsein getreten. Als Folge der Reform wurde aus Sicht vieler Lehrpersonen die (adaptive) Unterrichtsgestaltung nun nochmals anforderungsreicher, ohne dass die ihrer Meinung nach dafür notwendigen Ressourcen in genügendem Umfang bereitgestellt worden seien – beklagt wird etwa ein Mangel an Förderlektionen (Reusser et al., 2013).

Darüber hinaus wurde in der Heterogenitäts-Studie ermittelt, dass sich die Lehrpersonen unterscheiden haben in ihren *Überzeugungen*, inwieweit sie den gewachsenen Anforderungen genügen können: Einige gaben sich durchaus zuversichtlich, andere hingegen waren weniger optimistisch. Basierend auf bisherigen Forschungsbefunden ist anzunehmen, dass solche berufsbezogenen Überzeugungen von Lehrpersonen in Wechselwirkung mit weiteren Komponenten der professionellen Kompetenz eine gewisse Handlungsrelevanz entfalten, sodass sie als eine von vielen Bedingungen von Unterrichtsqualität angesehen werden können (Reusser, Pauli & Elmer, 2011). Angesichts dieser Bedeutung wurden im Rahmen des ersten Beitrags der Dissertation (Eckstein et al., 2013) Bedingungen integrationsbezogener Überzeugungen der an der Heterogenitäts-Studie teilnehmenden Lehrpersonen mit folgender *Fragestellung* analysiert:

Welchen Einfluss haben unterrichtsbezogene und strukturelle Faktoren auf den Optimismus von Lehrpersonen bezüglich der Umsetzung der integrativen Volksschule?

2.1.3. Bedingungen optimistischer Überzeugungen von Lehrpersonen

Integrationsbezogene Überzeugungen wurden in der früheren Forschung oftmals als eher handlungsferne Konstrukte operationalisiert. Die Lehrpersonen wurden z.B. nach ihrer Meinung gefragt, ob es für die Schüler/innen mit sonderpädagogischem Förderbedarf günstiger sei, in der Regelschule oder in einer Sonderklasse unterrichtet zu werden. Basierend auf solchen Fragen wurde wiederholt gefunden, dass Lehrpersonen die Integration gegenüber der Separation eher bevorzugen (Gebhardt et al., 2011; Kunz et al., 2010). Allerdings sind solche Einschätzungen von Lehrpersonen eher indirekt mit ihrem alltäglichen Handeln im Unterricht verbunden.

Demgegenüber standen in Beitrag I der Dissertation (Eckstein et al., 2013) *handlungsnahe Überzeugungen* von Lehrpersonen (Reusser et al., 2011) im Vordergrund. Dazu wurde das Konstrukt "*Integrationsoptimismus*" gebildet: Lehrpersonen mit hohen Werten auf dieser Skala sind beispielsweise zuversichtlich, dass sie mit ihrer Klasse die Lehrplanziele auch im Rahmen der Integration erreichen können, und dass dies keine Überforderung darstellt. Mit der in Beitrag I vorgestellten Analyse ging ich der Frage nach, was Lehrpersonen in dieser Hinsicht optimistisch macht.

Dazu habe ich die Angaben der an der Online-Befragung des Heterogenitäts-Projekts teilnehmenden 185 Klassenlehrpersonen anhand eines Strukturgleichungsmodells (SEM) untersucht, welches in Abbildung 1 dargestellt ist. Die Modellschätzung passt gut zu den Daten ($\chi^2=73.42$, $df=69$, n.s.; RMSEA=0.02; CFI=0.99). Ermittelt wurde der Einfluss von vier Prädiktoren auf das Konstrukt Integrationsoptimismus (IntOp): Erweiterte Didaktische Unterrichtspraxis (EDUP); Quote von Schülerinnen und Schülern mit Verhaltensauffälligkeiten pro Klasse (VA-Quote); Dienstalter; Klassengröße.

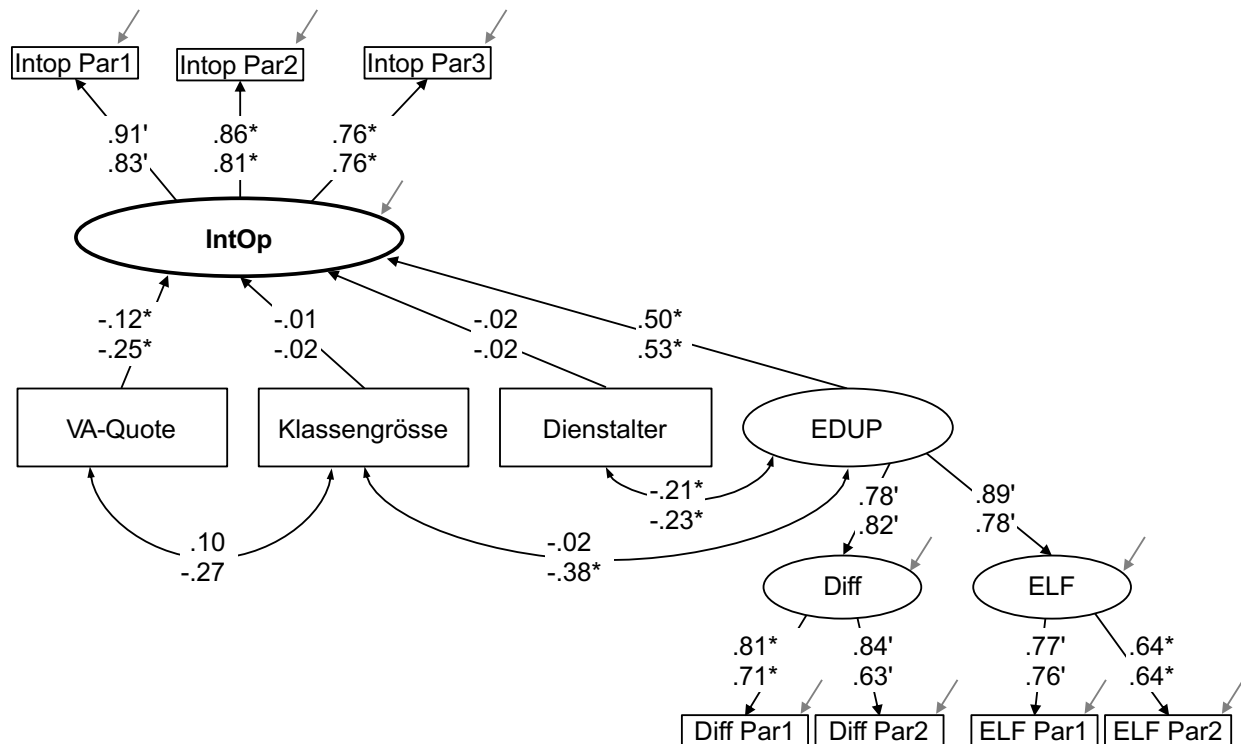


Abbildung 1. Prädiktion von Integrationsoptimismus im SEM (Eckstein et al., 2013, S. 12).

Anmerkungen: Alle im Manteltext der Dissertation abgebildeten SEM sind angelehnt an die Konventionen von Muthén und Muthén (2017). Abgebildet sind jeweils die standardisierten Parameter; mit Asterisk (*) gekennzeichnete Werte sind signifikant mit einer maximalen Irrtumswahrscheinlichkeit von $p < .05$. In Abbildung 1 bedeutet die Kennzeichnung mit Apostroph ('), dass der Parameter aus Identifikationsgründen fixiert wurde (keine statistische Absicherung); der obere Wert gilt jeweils für Primarlehrpersonen ($n=92$), der untere Wert für Sekundarlehrpersonen ($n=93$).

Die Klassengröße und das Dienstalter haben sich nicht als signifikante Einflussgrößen auf den Integrationsoptimismus erwiesen. Diese Nulleffekte haben weiterführende Fragen nach nicht-linearen Zusammenhängen bzw. nach kritischen Grenzwerten aufgeworfen, die aber in der vorliegenden Arbeit nicht weiter verfolgt werden. Von den zwei anderen Prädiktoren gehen signifikante Effekte aus:

- (1) Der bedeutsamere Prädiktor ist eine *Erweiterte Didaktische Unterrichtspraxis* (EDUP)
 - Lehrpersonen, die angaben, regelmässig binnendifferenziert und auf der Grundlage

Erweiterter Lehr-Lernformen zu unterrichten, äussern sich vergleichsweise optimistisch im Hinblick auf die Realisierbarkeit der Integration ($\beta_{Primar} = .50^*$; $\beta_{Sek} = .53^*$).

- (2) Der zweite signifikante Prädiktor ist die *Quote von Schülerinnen und Schülern mit Verhaltensauffälligkeiten pro Klasse* (VA-Quote), welche den Integrationsoptimismus der Lehrpersonen negativ beeinflusst – je höher diese Quote in ihrer eigenen Klasse, desto weniger optimistisch sind die Lehrpersonen, dass Integration gelingen kann ($\beta_{Primar} = -.12^*$; $\beta_{Sek} = -.25^*$).

Ausgangspunkt für die nachfolgenden Überlegungen und Analysen bildet der *negative Einfluss der VA-Quote*. Die modellierte Kausalrichtung dieses Effekts impliziert folgende Interpretation: Aufgrund einer zunehmenden Quote von Schülerinnen und Schülern mit Verhaltensauffälligkeiten pro Klasse verringert sich der Optimismus der Lehrpersonen.

Da die Lehrpersonen nach eigenen Ermessen angegeben haben, wie viele Schüler/innen ihrer Klasse verhaltensauffällig sind, und die Analyse zudem auf Querschnittdaten beruht, lässt sich die *Kausalrichtung auch umgekehrt interpretieren*: Die integrationsbezogenen Überzeugungen der Lehrpersonen haben einen Einfluss darauf, wie viele Schüler/innen ihrer Klasse sie als verhaltensauffällig einschätzen. Diese alternative Deutung liesse sich *etikettierungstheoretisch* (Becker, 1963/1981) begründen. Es könnte zum Beispiel sein, dass Lehrpersonen bei zunehmendem Optimismus grenzwertiges Verhalten ihrer Schüler/innen nachsichtiger beurteilen und in der Folge zurückhaltender sind mit der Zuschreibung von Verhaltensauffälligkeit. Oder andersherum: Je skeptischer eine Lehrperson gegenüber der Integration eingestellt ist, desto eher empfindet sie bestimmte Verhaltensweisen als unangemessen und etikettiert entsprechend mehr Schüler/innen als verhaltensauffällig.

Es gibt auch noch eine dritte Deutungsmöglichkeit, welche die beiden gegensätzlichen Interpretationsversuche integriert: *Viele Lehrpersonen sind tendenziell skeptisch gegenüber der Integration von Schülerinnen und Schülern mit Verhaltensauffälligkeiten*. So interpretiert deckt sich der ermittelte Effekt mit Ergebnissen der im Heterogenitäts-Projekt ebenfalls durchgeführten qualitativen Fallstudien (Reusser et al., 2013) sowie mit Resultaten einer Masterarbeit, welche die in einem freien Format gegebenen Antworten der Lehrpersonen auf vier offen gestellte Fragen innerhalb der Online-Umfrage inhaltsanalytisch untersuchte (Pappa, 2012):

Nach Einschätzungen der Lehrpersonen nehmen vor allem verhaltensauffällige Kinder zu viel Raum ein in der Klasse, so dass sie den „Unterricht massiv belasten“ (S15_OS_RLp) und „die Lehrperson sehr viel Energie aufwenden muss, um diese Verhaltensauffälligen in Schach zu halten“ (S17_OS_SLp). „Es gibt Kinder, die eine Klasse und deren Lehrperson überfordern“ (S3_OS_RLp) gibt eine Oberstufenlehrperson zu bedenken. Diese Kinder behindern die Lehrpersonen, einen „konstruktiven Unterricht“ (S17_OS_SLp) zu gestalten und ‚bremsen‘ so die Klasse. Allgemein nehmen diese Kinder gemäss Aussagen der Lehrpersonen zu viel Aufmerksamkeit in Anspruch. (S. 83)

Diese Befunde des Heterogenitäts-Projekts passen zur Forschungsliteratur, wonach Kinder und Jugendliche mit *Verhaltensauffälligkeiten als die am schwierigsten zu integrierenden Schüler/innen* gelten (Bless et al., 2010; de Boer, Pijl & Minnaert, 2011; Gebhardt et al., 2011; Liesen & Luder, 2011; Meijer, 2003; Preuss-Lausitz & Textor, 2006). Dabei zeichnen sich nicht nur Lehrpersonen durch diese skeptische Haltung aus – sie ist im gesamten (fach-)öffentlichen Diskurs³ verbreitet, wie ich im folgenden Kapitel 2.1.4 darlegen werde.

³ Als "(fach-)öffentlichen Diskurs" bezeichne ich in dieser Arbeit eine empirisch nicht leicht fassbare gesellschaftliche Diskussion und Kommunikation zum Thema der (integrativen) Schule. In schriftlicher Form zeigt sich diese beispielsweise in Zeitungsartikeln, Leserbriefen oder Online-Foren. Am Diskurs beteiligen sich nebst den Medienschaffenden auch Lehrpersonen, Fachpersonen schulischer Heilpädagogik, Psychologinnen und Psychologen, Ärzte und Ärztinnen, Erziehungswissenschaftler/innen sowie weitere Personen in ihrer Rolle als Eltern, Mitglieder von Schulpflegen u.a..

2.1.4. Skepsis gegenüber Schülerinnen und Schülern mit Verhaltensauffälligkeiten

Im (*fach-*)*öffentlichen Diskurs* wird diskutiert, ob Integration bei Verhaltensauffälligkeit überhaupt gelingen kann, wobei diese Kontroverse nicht immer sachlich ausgetragen wird (Bernhardt, 2007). Die folgenden Medien-Schlagzeilen stammen aus renommierten Zeitungsorganen und illustrieren anekdotisch, wie pointiert die Skepsis zuweilen vorgetragen wird:

- Verhaltensauffällige Kinder: "Die Not ist riesengroß" (Spiewak, 2010)
- Du störst! (Spiewak, 2014)
- Der Bub tut nicht gut: Die Grenzen der Integration (Binswanger, 2013a)
- Nicht mehr zumutbar (Binswanger, 2013c)
- Integrieren, bis die Polizei kommt (Binswanger, 2013b)
- Sonderschüler treiben Lehrer ans Limit (Pastega, 2015)
- Schiffbruch einer schönen Idee (Donzé, 2016)
- Störenfriede raus! (Burri, 2017)

Diese Zeitungsartikel vermitteln den Eindruck einer schwierigen Gesamtsituation, was vermutlich in Wechselwirkung mit der Wahrnehmung der Volksschule durch die breite Öffentlichkeit steht. In dieser Hinsicht konnte Schwab (2015) zeigen, dass sich (nebst Lehrpersonen auch) viele Personen aus der Allgemeinbevölkerung Sorgen machen, was die Integration von Kindern und Jugendlichen mit Verhaltensauffälligkeiten anbelangt.

Allerdings gibt es die Diskussion um den Umgang mit herausfordernden Schülerinnen und Schülern nicht erst seit der bildungspolitischen Trendwende in Richtung stärkerer Integration bzw. Inklusion. Bei *historischer Betrachtung* des Diskurses (Deluigi, 2016; Göppel, 2010; Keller, 2014; Stechow, 2015) kann geschlussfolgert werden, dass viele der heute

thematisierten Probleme und Bedenken schon wesentlich früher – und damit vor der schulpolitischen Trendwende in Richtung Integration bzw. Inklusion – in ähnlicher Weise vorgetragen wurden.

Seit einigen Jahren mehren sich in zunehmendem Maße Klagen von Lehrern an Grund- und Hauptschulen über starke Leistungsschwächen und Verhaltensauffälligkeiten ihrer Schüler, über Störversuche im Unterricht, über Aggressivität gegenüber Lehrkräften, Mitschülern, Sachmitteln wie auch über völlige Passivität, Gleichgültigkeit oder Resignation [...] Berichte in Presse, Rundfunk und Fernsehen beunruhigen und verunsichern die am Bildungswesen interessierte Öffentlichkeit. (Glötzl, 1976, S. 8)

Bei einer vergleichenden Analyse von Literatur und vor allem von Medienbeiträgen fällt auf, dass entgegen einer (im Prinzip erwartbaren) einheitlichen Verwendung des Begriffs Verhaltensauffälligkeit sehr *unterschiedliche Erscheinungsformen* damit bezeichnet und problematisiert werden. In der Gegenüberstellung einzelner Beiträge zeigen sich nebst Übereinstimmungen denn auch klare Differenzen, Übergeneralisierungen und durchaus auch Widersprüche.

So ist kaum (generalisierend) davon auszugehen, dass "Passivität, Gleichgültigkeit oder Resignation" (Glötzl, 1976) tatsächlich zahlreiche Lehrpersonen "ans Limit" (Pastega, 2015) treiben. Und die "Grenzen der Integration" (Binswanger, 2013a) verorten wahrscheinlich die wenigsten Lehrpersonen bei den "Zerstreuten" (Stechow, 2015), die an einem Aufmerksamkeitsdefizit-Syndrom (ADS) leiden. Kommen zu diesem Syndrom jedoch Hyperaktivität und Impulsivität hinzu (ADHS), werden betroffene Schüler/innen schon eher als "Störenfriede" (Burri, 2017) gesehen.

Der (fach-)öffentliche Diskurs ist überdies durchaus auch *ambivalent*, etwa wenn die medikamentöse Therapie störender Schüler/innen mit Methylphenidat (z.B. Ritalin) diskutiert wird (Dörpinghaus, 2009; Göppel, 2014) – oder wenn grundsätzliche Fragen erörtert werden, die

sich darauf beziehen, welche Verhaltensweisen heutzutage noch als "normal" gelten und welche dagegen als "krank" erachtet, also pathologisiert werden (Bleich, 2013).

Als *Zwischenfazit* lässt sich festhalten, dass viele Lehrpersonen tendenziell skeptisch eingestellt sind gegenüber der integrativen Schulung von Kindern und Jugendlichen mit Verhaltensauffälligkeiten. Diese Skepsis ist im gesamten (fach-)öffentlichen Diskurs verbreitet und wird von den Medien zuweilen pointiert vorgetragen. Allerdings werden dabei unterschiedliche Erscheinungsformen von Verhaltensauffälligkeit problematisiert, sodass gewisse Diskursbeiträge im Quervergleich widersprüchlich anmuten. Deshalb ist nicht ganz klar, wogegen sich die von der Allgemeinheit geteilte Skepsis genau richtet bzw. welches Ausmass das damit in Zusammenhang gebrachte Problem von Lehrpersonen annimmt.

2.1.5. Schwierigkeiten bei der Ausmarchung des Problemausmasses

Um die Skepsis der Lehrpersonen besser einordnen zu können, habe ich recherchiert, wie viele Kinder und Jugendliche mit Verhaltensauffälligkeiten gegenwärtig integrativ an Regelschulen unterrichtet werden. Jedoch zog diese Recherche mehr Fragen nach sich, als sie klären konnte. Das Problem liegt daran, dass "Verhaltensauffälligkeit" selbst im rein akademischen Diskurs als *unscharfer Sammelbegriff* verwendet wird, welcher externalisierende und internalisierende Formen mit unterschiedlichem Schweregrad umfasst. Dieses facettenreiche Phänomen kann gar "nicht punktgenau definiert werden" (Fröhlich-Gildhoff, 2007, S. 21). Aufgrund dieser definitorischen Unschärfe wurde Verhaltensauffälligkeit auch in der bisherigen Forschung heterogen konzeptualisiert, sodass verfügbare Resultate immer mit Blick auf die der jeweiligen Studie zugrundeliegende Operationalisierung zu interpretieren sind.

Beispielsweise ermittelten Berg, Imhof, Kollera, Schmidt und Ulber (1998) anhand einer Befragung von 165 Lehrpersonen zu deren 3665 Schülerinnen und Schülern, dass bei *40.9 Prozent aller Jungen* der 4. Jahrgangsstufe "mindestens einmal ein bestimmtes Verhalten als "stark auffällig" eingestuft wurde" (S. 284). Um diese Zahl angemessen zu deuten bedarf es der Kenntnis über den Kriterienkatalog, welcher bei der Umfrage eingesetzt wurde: Das Spektrum umfasste externalisierende Formen wie "Aggressives Verhalten" sowie internalisierende Formen wie "Depressive Verstimmungen". Am häufigsten haben die Lehrpersonen ihre Schüler/innen aber in den *low-level Kategorien* "Unkonzentriertheit" und "Ungenauigkeit" als stark auffällig eingeschätzt (S. 283). Folglich ist davon auszugehen, dass die ermittelte Prävalenzrate von rund 41 Prozent nicht nur den Anteil "nicht mehr zumutbarer" (Binswanger, 2013c) "Störenfriede" (Burri, 2017) beziffert, sondern auch weit weniger problematische Fälle einschliesst.

Aufgrund der Mehrdeutigkeit des Begriffs Verhaltensauffälligkeit bevorzugen einige Wissenschaftler/innen den aus ihrer Sicht präziseren Begriff "*Verhaltensstörung*" (Ahrbeck & Willmann, 2010; Myschker & Stein, 2014). Dagegen könnte man einwenden, dass sich der Begriff Verhaltensauffälligkeit im (fach-)öffentlichen Diskurs durchgesetzt hat, weil er als wertneutral gilt, wodurch sich das Risiko verringert, die Betroffenen zu stigmatisieren. Die zitierten Autoren halten dagegen, dass mit dem aus ihrer Sicht euphemisierend verwendeten Begriff der Verhaltensauffälligkeit eine Stigmatisierung gegebenenfalls verschleiert, nicht jedoch verhindert werde. Der Störungsbegriff sei deshalb ehrlicher und zugleich präziser, weil er *low-level Kategorien* wie "Ungenauigkeit" ausschliesse. Dieser Argumentation ist aber entgegenzuhalten, dass auch Verhaltensstörung als Sammelbezeichnung mehrere Störungsbilder umfasst und nicht einheitlich verwendet wird. Darüber hinaus impliziert der Auffälligkeitsbegriff, dass gezeigtes Verhalten als auffällig gedeutet und bewertet wird. Diese

Argumente bilanzierend favorisiere auch ich den Auffälligkeitsbegriff, und verwende ihn mehrheitlich in der Dissertation. Dennoch bin ich bei meiner Literaturrecherche auch auf den Forschungsstand zu Verhaltensstörungen eingegangen. Die Ergebnisse dieser Recherche lege ich im folgenden Abschnitt kurz dar.

Verschiedene epidemiologische Studien zusammenfassend weisen ungefähr 6-10 Prozent aller Kinder und Jugendlichen eine behandlungsbedürftige psychopathologische Störung auf, weitere ca. 10-13 Prozent werden als Grenzfälle betrachtet (Hurrelmann, Klocke, Melzer & Ravens-Sieberer, 2003; Ihle & Esser, 2002; Schwenck, 2016). Diese Prävalenzraten beziehen sich auf die Gesamtheit aller psychischen Störungsbilder bei Kindern und Jugendlichen.

Bei exklusiver Betrachtung der Gefühls- und Verhaltensstörungen wird basierend auf den mittleren Prävalenzraten geschätzt, dass etwa 15 Prozent aller Schüler/innen einen *sonderpädagogischen Förderbedarf* aufweisen (inkl. Grenzfälle) (Dlugosch, 2010; Fröhlich-Gildhoff, 2007; Myschker & Stein, 2014). Aber auch diese Zahl beschreibt das interessierende Phänomen aus zwei Gründen eher schemenhaft: Erstens, die Resultate der zur Schätzung herangezogenen, einzelnen Studien divergieren teilweise beträchtlich. Zweitens, die Prävalenzraten umfassen auch internalisierende Störungen – doch einem depressiven Kind würde vermutlich keine Lehrperson vorwerfen: "Du störst!" (Spiewak, 2014).

Darüber hinaus bleibt unklar, wie viele Kinder und Jugendliche mit Verhaltensstörungen integrativ unterrichtet werden und wie viele in den weiterhin existierenden sonderpädagogischen Fördereinrichtungen separativ beschult werden. Denn auch die verfügbaren *amtlichen Bildungsstatistiken* tragen wenig zur Klärung dieser Frage bei. So scheinen beispielsweise offizielle Förderquoten in Deutschland das Problem aufgrund der bildungsökonomisch bedingten

Knappheit von Förderangeboten eher zu unterschätzen (Opp, 2009; Preuss-Lausitz, 2005; Wachtel, 2010; Willmann, 2010).

In der Schweiz liegen gemäss meiner Recherche keine verlässlichen Informationen auf Bundesebene vor. So finden sich weder im "Bildungsbericht Schweiz" (Annen et al., 2010) noch im Bericht "Bildungsmosaik Schweiz" des Bundesamtes für Statistik (Ryser & von Erlach, 2007) schweizweite Angaben zu unterschiedlichen Typen sonderpädagogischen Förderbedarfs, was unter anderem auf kantonal unterschiedliche Zuweisungspraxen zurückzuführen ist. Im Kanton Zürich beispielsweise können Schüler/innen mit besonderen pädagogischen Bedürfnissen von Fachpersonen schulischer Heilpädagogik im Rahmen der integrativen Förderung (IF) unterstützt werden, was nicht zwingend eine schulpsychologische Abklärung voraussetzt (Bildungsdirektion Kanton Zürich, 2007, 2011). Dieses niederschwellige Verfahren ist pädagogisch sinnvoll, führt aber dazu, dass die differenziellen Gründe der Zuweisung nicht standardisiert erfasst werden und daher nicht Eingang in eine amtliche Statistik finden.

Zwischenfazit: Weil das Phänomen "Verhaltensauffälligkeit" in der bisherigen Forschung verschiedenartig definiert und operationalisiert wurde, gestaltet sich die Befundlage zum effektiven Problemausmass als heterogen und lässt sich daher kaum generalisieren. Zudem fehlen verlässliche amtliche Statistiken, sodass insbesondere auch für die Schweiz unklar ist, wie viele Schüler/innen mit Verhaltensauffälligkeiten an der Regelschule unterrichtet werden.

2.1.6. Grenzen der individualpsychologischen Suchlinie

Die Heterogenität der Befundlage zur Prävalenz von Verhaltensauffälligkeiten lässt sich damit erklären, dass es kein allgemeingültiges Verständnis dazu gibt, welche Verhaltensweisen

als normal bzw. als auffällig zu bewerten sind. Deshalb lässt sich auch das effektive Problemausmass kaum als objektive Tatsache in einem ‘positivistischen’ Sinne ermitteln. Das bedeutet, dass das Phänomen Verhaltensauffälligkeit mit einem rein individualpsychologischen, personorientierten Fokus auf einzelne Schüler/innen *einseitig und unvollständig konzeptualisiert* wird.

Als Konsequenz gehe ich in den nachfolgenden Erkenntnisbemühungen meiner Dissertation von einem erweiterten Problemverständnis aus: Die Skepsis gegenüber Schülerinnen und Schülern mit Verhaltensauffälligkeiten rührt daher, dass sie vergleichsweise häufig Verhaltensweisen zeigen, welche die Lehrpersonen oder Mitschüler/innen in der Regel ablenkt, provoziert, verärgert, frustriert oder belastet. Daraus resultieren *Störungen des Unterrichts* (Lohmann, 2003; Nolting, 2012; Winkel, 2011), und diese stellen die eigentliche Herausforderung bei Verhaltensauffälligkeiten in der Regelschule dar. Personale (individualpsychologische) Eigenschaften einzelner Schüler/innen (z.B. ADHS) können zur Problematik beitragen, vermögen sie aber nicht monokausal zu erklären und stehen deshalb nicht mehr *allein* im Zentrum der Untersuchung. Um zu einer angemesseneren Erklärung zu kommen, braucht es die Berücksichtigung der Sichtweise der Rezipienten, der von einer ‘Störung’ Betroffenen.

Diese Neujustierung des Erkenntnisinteresses hat auch zur Konzeption des SUGUS-Projekts beigetragen. In diesem werden neben den objektiven Merkmalen störenden *Schülerverhaltens* (Produktion) ebenfalls das subjektive *Störungsempfinden* (Rezeption) von Lehrpersonen, Schülerinnen und Schülern als konstitutive Merkmalsdimensionen von Unterrichtsstörungen einbezogen und untersucht. In welchem Rahmen und vor welchem theoretischen Hintergrund dies geschieht, wird in den nachfolgenden Abschnitten ausgeführt.

2.2. Beitrag II: Gestörter Unterricht – Entflechtung zentraler Merkmalsdimensionen

Abstract aus der Originalarbeit

Unterrichtsstörungen gelten als eine der belastendsten Herausforderungen in der schulpädagogischen Praxis, was auch in der gegenwärtigen Inklusionsdebatte mit teilweise neuen Akzenten thematisiert wird. Dennoch fehlt es an profundem Wissen zum subjektiven Störungsempfinden von Lehrpersonen und Schülern sowie zur Bedeutung von Störungskontexten. Diese Forschungslücke aufgreifend wurde in der Studie zur Untersuchung gestörten Unterrichts (SUGUS) auf der Basis eines interaktionistischen Theorierahmens ein mehrperspektivisch angelegtes Instrumentarium zur Erfassung von Unterrichtsstörungen entwickelt und einem Pretest mit 11 Klassen des 5. Schuljahrs unterzogen. Präsentiert werden die faktorielle Struktur der Instrumente mittels eines Mehrebenen-Strukturgleichungsmodells und deskriptive Kennwerte zu den theoretischen Konstrukten. (Eckstein, Grob, et al., 2016, S. 113)⁴

Beitrag II meiner Dissertation entstand im Rahmen des Forschungsprojekts SUGUS, welches im nachfolgenden Kapitel 2.2.1 eingeführt wird. Die Kapitel 2.2.2 bis 2.2.6 behandeln den Theorierahmen des SUGUS-Projekts in einem ausführlichen Sammelreferat – in der Originalarbeit war dies nur in begrenztem Umfang möglich. In den Kapiteln 2.2.7 und 2.2.8 werden sodann die methodologischen Grundlagen der quantitativen Teilstudie des SUGUS-Projekts beschrieben. Dabei steht der SUGUS-Pretest im Vordergrund, der auch das Hauptthema von Beitrag II der Dissertation war. In den Kapiteln 2.2.9 und 2.2.10 werden die Befunde des Pretests vorgestellt. Zudem wird dargelegt, wie die Ergebnisse den teilnehmenden Lehrpersonen zurückgemeldet wurden. Zentrale Schlussfolgerungen und offen gebliebene Fragen werden in Kapitel 2.2.11 im Hinblick auf die SUGUS-Hauptstudie diskutiert.

⁴ Die vollständigen bibliographischen Angaben der Originalarbeit lauten:
Eckstein, B., Grob, U. & Reusser, K. (2016). Unterrichtliche Devianz und subjektives Störungsempfinden. Entwicklung eines Instrumentariums zur Erfassung von Unterrichtsstörungen. *Empirische Pädagogik*, 30(1), 113-129.

2.2.1. Das SUGUS-Projekt

Die „*Studie zur Untersuchung gestörten Unterrichts – SUGUS*“ ist ein vom Schweizerischen Nationalfonds gefördertes Forschungsprojekt (Projekt-Nr.: 100019_152722; Laufzeit: 2014 - 2019), welches von Prof. em. Dr. Kurt Reusser geleitet wird. Dem Projekt liegt ein *interaktionistisches Problemverständnis* zugrunde: Unterrichtsstörungen können aus unangemessenen Verhaltensweisen von Schülerinnen, Schülern oder Lehrpersonen hervorgehen oder durch externe Quellen verursacht werden (Menzel, 2009; Montuoro & Lewis, 2015; Wicki & Kappeler, 2007). Um eine Störung handelt es sich dann, wenn in der Folge mindestens eine im Unterricht involvierte Person vom kognitiv-inhaltlichen Sinnfluss abgelenkt wird oder wenn ihre affektiv-emotionale Befindlichkeit beeinträchtigt wird (Eckstein, Grob, et al., 2016).

Demnach bestehen Unterrichtsstörungen aus einem *objektiven Kern (Produktion)*, welcher in der jeweiligen Perspektive der beteiligten Akteure *subjektiv wahrgenommen, gedeutet und bewertet wird (Rezeption)*. Dieser interaktionale Problemzusammenhang wird in der SUGUS-Studie auf unangemessenes Schülerverhalten sowie auf das darauf bezogene Störungsempfinden von Lehrpersonen, Mitschülerinnen und Mitschülern eingegrenzt. Von Lehrpersonen ausgehendes Fehlverhalten (Krumm, 2003; Krumm & Weiß, 2006) oder pädagogisch-didaktisches Missgeschick werden als Kontextbedingungen berücksichtigt, stehen aber nicht im Vordergrund der Untersuchung.

Unter dem Aspekt der *Produktion* gestörten Unterrichts interessiert demnach Schülerverhalten, welches von im Prinzip geltenden oder etablierten unterrichtlichen Normen abweicht (unterrichtliche Devianz; vgl. Kap. 2.2.2, 2.2.3). In der vorliegenden Dissertation wird dazu der Versuch unternommen, die Auftretenshäufigkeit devianten Schülerverhaltens möglichst objektiv zu bestimmen, sowie Differenzen in der Beschreibung unterrichtlicher Devianz durch

Lehrpersonen und Schüler/innen sichtbar zu machen bzw. in den statistischen Analysen zu kontrollieren.

Komplementär dazu interessiert unter dem Aspekt der *Rezeption* gestörten Unterrichts das Störungsempfinden der beteiligten Akteure im Sinne einer subjektiven Erlebensqualität (vgl. Kap. 2.2.4, 2.2.5). Dazu werden die Selbstwahrnehmung der Schüler/innen sowie die Fremdwahrnehmung der Lehrpersonen und Mitschüler/innen explizit erfasst und in die Analysen miteinbezogen.

Ein weiteres Ziel des SUGUS-Projekts besteht darin, Bedingungen beider Elemente (Devianz, Störungsempfinden) zu identifizieren. Die folgenden drei *Hauptfragestellungen* bringen das Erkenntnisinteresse der SUGUS-Studie zusammenfassend zum Ausdruck:

1. *Unter welchen Bedingungen des Unterrichts zeigen welche Lernenden welche Formen und Grade normabweichenden Verhaltens?*
2. *Unter welchen Bedingungen des Unterrichts wird normabweichendes Verhalten von Lehrpersonen, Schülerinnen und Schülern als wie stark störend erlebt?*
3. *Welche pädagogisch-didaktischen Handlungsmöglichkeiten sehen die Lehrpersonen?*

Diesen Fragestellungen geht das SUGUS-Projekt in zwei Teilstudien nach: Die *qualitative Teilstudie* beruht auf leitfadengestützten Interviews mit Lehrpersonen, Schülerinnen und Schülern (Hofstetter & Reusser, 2018; Marusic-Würscher & Reusser, 2018). Für die *quantitative Teilstudie* wurden die Sichtweisen von Lehrpersonen, Schülerinnen und Schülern mittels eines multiperspektivisch angelegten, standardisierten Fragebogens in 5. Klassen der Primarschule untersucht.

Die drei Originalarbeiten II bis IV der Dissertation entstanden im Rahmen der quantitativen SUGUS-Teilstudie und widmen sich bestimmten *Teilaspekten der ersten zwei Hauptfragestellungen* des Projekts. In diesem Kapitel 2.2 wird zunächst der Theorierahmen dargelegt, auf welchem das Projekt beruht. Diese Ausführungen beginnen im nachfolgenden Kapitel 2.2.2 mit einer Übersicht über unterrichtliche Normen, von denen das interessierende Schülerverhalten abweichen kann.

2.2.2. Schülerverhalten im Spiegel unterrichtlicher Normen

Mit dem postulierten "objektiven Kern" von Unterrichtsstörungen wird gleichsam unterstellt, dass bestimmtes Schülerverhalten aufgrund dessen *inhärenten Störungspotenzials* die Produktivität eines Lehr-Lernprozesses und/oder die Annehmlichkeit der damit verbundenen affektiv-emotionalen Lernatmosphäre gefährden kann. Im Folgenden wird argumentiert, dass es sich hierbei um Verhaltensweisen handelt, welche von im Prinzip geltenden unterrichtlichen Normen abweichen (Lamnek, 2007).

Unterrichtliche Normen existieren einerseits in Form *expliziter Regeln*, welche etwa im Rahmen des Classroom Managements mit der ganzen Klasse (Emmer & Evertson, 2013; Ophardt & Thiel, 2013; Stein & Stein, 2014) oder bei der Förderplanung mit einzelnen Kindern (Frölich, Döpfner & Banaschewski, 2014; Gentz, 2012; Hennemann, Hövel, Casale, Hagen & Fitting-Dahlmann, 2015) ausgehandelt werden. Andererseits prägen viele *implizite Verhaltenserwartungen* den Unterricht (Doyle, 2006), sodass die Bandbreite der im Unterricht geltenden Normen von ungeschriebenen, sittsamen, kulturellen Gepflogenheiten bis zu strafrechtlich relevanten Gesetzen reicht.

Ausgehend von diesen Vorüberlegungen werden *zwei Hauptkategorien unterrichtlicher Normen* unterschieden, welche die Schüler/innen im Unterricht beachten sollten: (1) Universell gültige soziomoralische Normen; (2) unterrichtsspezifische Normen.

(1) Universell gültige soziomoralische Normen

Wie in allen gesellschaftlichen Bereichen gilt auch im Unterricht eine Art elementarer, *soziomoralischer Verhaltenskodex* (Nunner-Winkler, 2012; Stein, 2014). Dazu gehören insbesondere die gegenseitige Achtung der physischen und psychischen Integrität aller beteiligten Personen sowie ein Mindestmass an Sorgfalt beim Umgang mit der materiellen Umwelt. Man soll einander *respektvoll* begegnen, was beispielsweise darin Ausdruck findet, dass sich die Schüler/innen und die Lehrperson vielerorts zur Begrüssung und zur Verabschiedung die Hand reichen. In Gesprächen gilt, dass man *einander zuhört und ausreden lässt*, auch in Konflikten. Unstimmigkeiten sollen *sozial verträglich* gelöst werden, es gilt Kompromisse zu finden. Geächtet wird dagegen der expressive Ausdruck von Aggression. Für die (unmündigen) Schüler/innen gilt im Besonderen, dass sie *Anweisungen der (erwachsenen) Lehrperson befolgen* sollen. Solche soziomoralischen Normen sind in den meisten Schweizer Lehrplänen enthalten. Grob und Maag Merki (2001) haben inhaltsanalytisch herausgearbeitet, welche Normen darin formuliert werden, es sind u.a.: Gemeinschaftsfähigkeit, Toleranz, Wertschätzung, Dialogfähigkeit, Kooperationsfähigkeit, Verantwortung für Mitmenschen.

Die Einhaltung dieses allgemeingültigen soziomoralischen Kodex ist eine *notwendige, nicht aber hinreichende Bedingung* für das Funktionieren unterrichtlicher Lehr-Lernprozesse. Dazu bedarf es zusätzlich der Einhaltung unterrichtsspezifischer Verhaltenserwartungen.

(2) Unterrichtsspezifische Normen

Unterrichtsspezifische Normen zeichnen sich dadurch aus, dass sie fast ausschliesslich im Klassenzimmer gelten. Ein typisches Beispiel im Deutschschweizer Unterricht besteht darin, dass die mündliche Kommunikation ab der Primarschulstufe in der *Standardsprache* erfolgen soll – schweizerdeutsche Wortmeldungen stellen somit meistens eine Normabweichung dar. Ähnlich ist es im Fremdsprachenunterricht, in welchem man sich vorrangig in der zu lernenden Zielsprache unterhalten soll. Solche Normen können *von Lektion zu Lektion, oder sogar situativ variieren*, bestimmte Settings können auch gegensätzliche Verhaltensanforderungen implizieren. So ist etwa in einer Sportlektion körperliche Bewegung ein curricular vorgegebenes Ziel während in einer Mathematikprüfung motorische Ruhe eine disziplinarische Grundvoraussetzung darstellt. Ein Moderator unterrichtsspezifischer Normen ist die *Sozialform* (Peterssen, 1982): Der lehrpersonenzentrierte Frontalunterricht (z.B. Klassengespräch, Lehrpersonenvortrag) beruht auf einer spezifischen Kommunikationskultur, in welcher *strenge Gesprächsregeln* gelten (Zevenbergen, 2001), darunter die Pflicht des Meldens (Aufstrecken). Zudem ist die *Bewegungsfreiheit stark eingeschränkt* – alle Schüler/innen sollen ruhig an ihrem Platz sitzen. Ähnlich restriktive Verhaltenserwartungen müssen auch während Phasen der stillen Einzelarbeit eingehalten werden, damit produktives Lernen möglich ist. Im Unterschied dazu gelten während Partner- oder Gruppenarbeitsphasen andere unterrichtsspezifische Normen. Das Ziel ist hier stärker der *kooperative und kokonstruktive Austausch* unter den Lernenden und weniger das Ideal des schweigsamen, bewegungsarmen individuellen Lernens. In einem solchen Unterricht dürften der *Geräuschpegel und die Bewegungsfrequenz* vermutlich etwas höher sein, dafür wird von den Schülerinnen und Schülern erwartet, dass sie Interaktionsnormen einhalten und auch Verantwortung für die Teamarbeit und für ihre eigenen Lernprozesse übernehmen.

Die Gesamtheit aller *allgemeingültigen (sozio-moralischen) und unterrichtsspezifischen Normen* regelt die sozialen Interaktionen in der Klasse sowie darauf aufbauend die unterrichtlichen Arbeitsprozesse. Dieser Rahmen unterrichtlicher Verhaltenserwartungen wird in der Dissertation als *kriteriale Bezugsnorm* herangezogen, um Schülerverhalten als normkonform oder normabweichend (deviant) zu beurteilen. Eine günstige Lernumgebung ist in der Regel dadurch gekennzeichnet, dass sich die Schüler/innen (und die Lehrpersonen) normkonform verhalten, während die Produktivität der Lehr-Lernprozesse und/oder die Annehmlichkeit der affektiv-emotionale Lernatmosphäre durch Normabweichungen potenziell gefährdet wird.

Deviantes Schülerverhalten wird in der Literatur unterschiedlich bezeichnet und verschiedenartig systematisiert (Pfitzner & Schoppek, 2000; Wettstein, 2008). Im SUGUS-Projekt wird die Bezeichnung "*unterrichtliche Devianz*" verwendet. Es existieren verschiedene Formen, die sich in Bezug auf *Schweregrad* und *Auftretenshäufigkeit* unterscheiden. Eine Systematisierung lässt sich aus der Befundlage der bisherigen Forschung ableiten: Gravierende Formen unterrichtlicher Devianz wie physische Gewalt kommen sehr selten vor. Vergleichsweise bagatellhafte Devianzformen wie Schwatzen ereignen sich häufiger (Beaman, Wheldall & Kemp, 2007; Berg et al., 1998; Crawshaw, 2015; Harrison, Vannest, Davis & Reynolds, 2012; Kyriacou, Avramidis, Hoie, Stephens & Hultgren, 2007; Müller, Fleischli & Hofmann, 2013).

Ausgehend von dieser Befundlage werden im SUGUS-Projekt *zwei Hauptkategorien* unterrichtlicher Devianz unterschieden: (i) Disziplinprobleme, (ii) dissoziales Verhalten. Diesen Kategorien werden in der nachfolgenden Liste konkrete Verhaltensweisen zugeordnet, um die interessierenden Phänomene zu veranschaulichen. Diese Liste ist weder vollständig, noch umfasst sie trennscharfe Unterkategorien.

(i) Disziplinprobleme

- *Akustische Unruhe*: pfeifen, summen, johlen, grölen, rufen, mit Dingen scheppern, klappern oder trommeln, Zwischengespräche führen
- *Motorische Unruhe*: herumzappeln, kippen, den Sitzplatz verlassen – obwohl alle Schüler/innen ruhig an ihrem Platz sitzen müssten
- *Oppositionalität*: Anweisungen der Lehrperson verweigern, Wortmeldungen ohne Handaufhalten, freche Antworten, nicht-aufgabenbezogene Aktivitäten, Dinge umherwerfen

(ii) Dissozialität

- *Distanzlosigkeit*: andere Kinder anpusten, betatschen, anstupsen, kitzeln, erschrecken, beim Anstehen vordrängeln
- *Offene Aggression*: andere Kinder auslachen, beleidigen; Dinge wegnehmen, umwerfen, kaputt machen; hitzig streiten, drohen, festhalten, boxen, treten
- *Verdeckte Aggression*: gemein reden über andere Kinder hinter deren Rücken, Gerüchte verbreiten, andere ausschliessen, heimlich Dinge besudeln, verstecken

Dieser Katalog steht für den postulierten objektiven Kerns von Unterrichtsstörungen. Die Verhaltensweisen stellen Abweichungen allgemeingültiger oder unterrichtsspezifischer Normen dar, weswegen sie ein inhärentes Störungspotenzial aufweisen: Sie können die Produktivität der kognitiven Lehr-Lernprozesse sowie die affektiv-emotionale Befindlichkeit der Lehrperson oder der Mitschüler/innen gefährden. Aus pädagogisch-diagnostischer Sicht stellt sich die Frage, wie solches Verhalten zustande kommt und wie es sich verhindern oder eindämmen lässt, worauf ich im folgenden Kapitel 2.2.3 eingehen werde.

2.2.3. Unterrichtliche Devianz: Ätiologie, Prävention, Intervention

Manche Schüler/innen zeigen im Unterricht häufiger deviantes Verhalten als andere, weshalb viele Studien *ontogenetisch bedingte Verhaltensdispositionen* als Ursachen unterrichtlicher Devianz erforschen. Dazu gehören die bereits erwähnten Gefühls- und Verhaltensstörungen im Allgemeinen (Bakken, Obiakor & Rotatori, 2012; Reid, Gonzalez, Nordness, Trout & Epstein, 2004) oder das Aufmerksamkeitsdefizit-Hyperaktivitäts-Syndrom (ADHS) im Besonderen (Carroll et al., 2006; Kuester & Zentall, 2012; Tuvblad, Zheng, Raine & Baker, 2009). Viele solcher Studien beruhen auf der theoretischen Grundlage des *Bio-Pscho-Sozialmodells* (Petermann & Natzke, 2008), welches nicht nur Individualmerkmale der betroffenen Kinder und Jugendlichen thematisiert, sondern beispielsweise auch Merkmale ihrer Sozialisation (z.B. autoritärer Erziehungsstil). Allerdings werden diese ursächlichen Sozialisationseinflüsse oftmals in ihrer inkorporierten Form beim Individuum (z.B. geringe Sozialkompetenz) als Bedingungen unterrichtlicher Devianz erforscht.

Demgegenüber untersucht die bildungssoziologische Forschung makrosystemische Zusammenhänge als Ursache unterrichtlicher Devianz. Beispielsweise hat Zevenbergen (2001) auf der Grundlage von *Bourdieu's Habitustheorie* (1983) herausgearbeitet, dass Schüler/innen aus bildungsnahen Familien gewisse unterrichtsspezifische Verhaltenserwartungen bereits während ihrer Primärsozialisation verinnerlicht haben, z.B. implizite Regeln des fragend-entwickelnden Lehrgesprächs. Diese Chance haben Kinder in bildungsfernen Familien nicht, sodass sie strukturell bedingt ein höheres Risiko aufweisen, sich im Unterricht normabweichend zu verhalten. Aufgrund solcher Befunde fordert Christe (2010) pädagogische Konzepte, die auch

"nicht-mittelschichtspezifische Denk- und Lernformen akzeptiert" (S. 48), z.B. weniger Kopflastigkeit und stärkere Gewichtung des praktischen, erfahrungsbasierten Lernens⁵.

Ebenfalls zur Familie der soziologischen Erklärungsansätze gehört *Mertons Anomietheorie* (1949/1995), wonach unterrichtliche Devianz als Resultat einer Ziel-Mittel-Diskrepanz aufgefasst werden kann: Fehlen den Schülerinnen und Schülern legitime Mittel (z.B. bildungsnaher Habitus; intellektuelle Kompetenzen) um sozial anerkannte Ziele zu erreichen (z.B. Sicherung eines hohen sozialen Status innerhalb der Klasse, Beachtung durch die Lehrperson), setzen sie alternative Mittel zur Zielerreichung ein. Nach diesem Verständnis kann unterrichtliche Devianz als dysfunktionale Copingstrategie aufgefasst werden (Stein, 2015).

Gemäss diesen individualpsychologischen und soziologischen Ansätzen wird aus unterschiedlichen Blickwinkeln geschlussfolgert, dass gewisse Kinder ihre Schullaufbahn bereits unter nachteiligen Voraussetzungen starten und deshalb mehr Mühe haben als andere Kinder, sich gemäss den im Unterricht geltenden Normen zu verhalten. Dabei werden *distale Einflussgrössen* des unterrichtlichen Verhaltens hervorgehoben und erforscht (z.B. genetische Prädisposition, ungünstige familiäre Verhältnisse, makrostrukturell bedingte Nachteile).

Eine alternative Theorie- und Forschungstradition untersucht dagegen *proximale Bedingungen des aktuellen Verhaltens*, z.B. den situativen Kontext (Cialdini & Reno, 1990). Dieser Zugang ist aus erziehungswissenschaftlicher Perspektive von besonderem Interesse, weil sich der unterrichtliche Kontext im Hinblick auf Verhaltenssteuerung in verschiedener Hinsicht gestalten lässt (Gest & Rodkin, 2011), während ontogenetisch bedingte Verhaltensdispositionen

⁵ Diese Schlussfolgerungen erinnern an Deweys (1916/1994) Konzept der "Erziehung durch und für Erfahrung", worauf in der vorliegenden Arbeit aber nicht näher eingegangen werden kann.

wie auch makrosystemische Gegebenheiten kaum beeinflussbar sind. Damit verschiebt sich der Fokus auf die *Aktualgenese unterrichtlicher Devianz* (Stein & Stein, 2014).

Die Unterrichtsforschung hat eine Reihe *pädagogisch-didaktischer Risikofaktoren* identifiziert, unter denen sich Schüler/innen vermehrt normabweichend verhalten: Mangelhafte Strukturierung des Unterrichts, Unter- und Überforderung der Schüler/innen (Schönbächler et al., 2009); langweilige, monotone Unterrichtssequenzen (Breidenstein, 2006); ungünstige Beziehungsmuster innerhalb der Klasse bzw. zur Lehrperson (Eder, 2006; Wettstein et al., 2018); lange und häufige Wechselphasen zwischen verschiedenen Unterrichtseinheiten (Wettstein, Thommen & Eggert, 2010); unangemessene Formen der Disziplinierung wie öffentliches Blossstellen, unablässiges Massregeln oder aversives Strafen (Hempel-Jorgensen, 2009; Holtappels, 2000; Krumm, 2003; Krumm & Weiß, 2006).

Die Forschung hat aber auch *Gelingensbedingungen* einer wirksamen Störungsprävention sowie effiziente Interventionsstrategien identifiziert. Zu den am besten untersuchten Merkmalen des Unterrichts gehört das *Classroom Management* (Emmer & Sabornie, 2015), wobei sich das Verständnis der Klassenführung⁶ seit Kounins Klassiker (1976/2006) stark weiterentwickelt hat, wie folgende online verfügbare Begriffsdefinition beispielgebend⁷ veranschaulicht:

Klassenführung ist alles, was Lehrpersonen durch ihre Aktivitäten und Haltungen zur Steuerung der Interaktionen in der Klasse beitragen. Der Begriff Klassenführung wird hier weit gefasst. Eine gute Strukturierung des Unterrichts und eine effiziente Organisation des Unterrichts führen zu guter Nutzung der Lernzeit und beugen Disziplinstörungen vor. Auf Disziplinprobleme reagieren die Lehrpersonen flexibel und situationsangemessen. Die Lehrperson sorgt auch für ein gutes Klima innerhalb der Klasse. Sie unterstützt die Schülerinnen und Schüler darin, selbst Verantwortung für die Klassengemeinschaft zu übernehmen. (Bildungsdirektion Kanton Zürich, 2018)

⁶ Classroom Management und Klassenführung werden in dieser Arbeit synonym verwendet.

⁷ Helmke (2009) und Seidel (2009) stützen sich auf diese Definition.

Im SUGUS-Projekt gehen wir auf der Basis unseres projektbezogenen Theorierahmens (Hofstetter & Reusser, 2016) davon aus, dass das theoretische Konstrukt "Klassenführung" gemäss zeitgenössischem Verständnis *mehrere Dimensionen* umfasst und mittels *verschiedener Strategien* realisiert werden kann: "Es sind dies Routinen, Regeln und Rituale, Raumgestaltung, Lernunterstützung, die Vereinbarung von Lern- und Verhaltenszielen, Self Management sowie Peer Tutoring" (S. 27). Darüber hinaus bestehen wechselwirksame Zusammenhänge zwischen der Klassenführung und weiteren Merkmalen der Unterrichtsqualität. Dazu gehören namentlich die "inhaltliche Klarheit und Strukturiertheit" des Unterrichts sowie – in Verbindung mit Massnahmen der Beziehungsförderung (Mayr, 2006) – auch das "lernförderliche Klima" (Helmke, 2009; Seidel, 2009).

Eine erfolgreiche Klassenführung erfordert, dass die zu ergreifenden Massnahmen *dem didaktischen Setting angepasst* werden. Beispielsweise sind bei schülerzentrierten Lernumgebungen andere Strategien gefragt als beim lehrerzentrierten Klassenunterricht, womit auch unterschiedliche Rollen der Lehrpersonen einhergehen (Pauli & Reusser, 2000; Reusser, 1999). Aus praktischer Sicht stellt sich die Frage, wie die Schüler/innen dazu angehalten werden können, sich kontinuierlich normkonform zu verhalten, wenn sie die Lehrperson nicht konstant überwacht und das Lernhandeln nicht kleinschrittig strukturiert ist. Ein zentraler Lösungsansatz besteht darin, den Schülerinnen und Schülern vermehrt (*Eigen-Verantwortung für das Lernverhalten*) zu übertragen, was allerdings keine triviale Aufgabe darstellt (Haag & Streber, 2012; Ophardt & Thiel, 2013; Wannack & Herger, 2014).

In Anlehnung an den *Response to Intervention Ansatz* (RTI) (Hartke, Diehl, Mahlau & Voss, 2014) lässt sich Klassenführung nochmals weiter fassen⁸. Angewendet auf Interventionsmöglichkeiten bei unterrichtlicher Devianz bedeutet dies, die zu ergreifenden Massnahmen an den Schweregrad der Normverletzung anzupassen. Dabei umfasst der im Idealfall stufenlos regulierbare Handlungsspielraum elementare Methoden der allgemeinen Pädagogik und Didaktik bis zu sonderpädagogischen Einzelfall-Massnahmen, welche auf elaborierten Verlaufsdagnosen beruhen (Casale, Grosche, Volpe & Hennemann, 2017):

- Ignorieren
- Subtile pädagogische Massnahmen "didaktischer Art" (z.B. betont stoffbezogene Unterstützung von Lernenden zur Verschleierung erzieherischer Absichten)
- Explizite pädagogische Massnahmen (z.B. Lob, Ermahnung, Sanktion, Wiedergutmachung, Time-Out innerhalb der Schule)
- Niederschwellige sonderpädagogische Massnahmen (z.B. integrative Förderung im Rahmen der regulären Klasse)
- Weitreichende sonderpädagogische Massnahmen (z.B. schulpsychologische Abklärung, diagnosebasierte Förderplanung, Beschulung in einer externen Förder-/ Sonderschule)

⁸ In der gegenwärtigen Fachliteratur werden die Vor- und Nachteile unterschiedlicher Konzepte einer inklusiven Didaktik diskutiert, wobei der RTI-Ansatz teilweise auch kritisiert wird (Moser Opitz, 2014). Auf diese Diskussion kann in der vorliegenden Arbeit aber nicht näher eingegangen werden.

Lehrpersonen haben somit vielfältige *proaktive und reaktive Handlungsoptionen*, um Unterrichtsstörungen im Rahmen der Klassenführung wirksam zu begegnen. Verfügbare empirische Befunde legen nahe: Die Grundsicht eines störungsarmen Unterrichts ist dessen angemessene Strukturierung durch die Lehrperson auf allen Ebenen (inhaltlich, organisatorisch, sozial) in Kombination mit einem emotional warmen Lernklima und autonomiefördernden Massnahmen (Jang, Reeve & Deci, 2010; Makarova, Schönbächler & Herzog, 2009; Pietsch, 2010, 2013).

Wenn die Grundvoraussetzungen eines strukturierten und motivierenden Unterrichts bei einem positiv erlebten Sozialklima gegeben sind, können darüber hinaus *binnendifferenzierende Massnahmen* störungspräventiv wirken (Borders, Bock & Michalak, 2012; Godwin et al., 2016; Tomlinson, 2014; Wettstein et al., 2010). Diese eröffnen den Schülerinnen und Schülern simultan mehrere Zugänge und Bearbeitungsmöglichkeiten des Stoffs, z.B. anhand verschiedener Leistungsniveaus (Klieme & Warwas, 2011; Niggli, 2013). Dies schmälert das Risiko, einzelne Lernende zu über- oder unterfordern, womit ein Schlüsselfaktor der Aktualgenese unterrichtlicher Devianz (Schönbächler et al., 2009) ausgehebelt wird.

Bei der Binnendifferenzierung handelt es sich um eine Qualitätsdimension der lehr- lernpsychologischen Tiefenstruktur des Unterrichts (Reusser, 2008), welche an der didaktischen Oberfläche verschiedenartig umgesetzt werden kann – etwa im Rahmen *offenen Unterrichts* (Bohl & Kucharz, 2010; Grünke, 2008; Kreuzberger, 2002; Lipowsky, 2002) oder durch die Inszenierung *Erweiterter Lehr-Lernformen* (Crocì, Imgrüth, Landwehr & Spring, 1995; Pauli, Reusser & Grob, 2010). Diese didaktischen Settings ermöglichen vermehrt individuelle Lern- und Denkwege, erweiterte Mitgestaltungsmöglichkeiten durch die Schüler/innen und kooperative Aktivitäten unter den Lernenden. Die Forschung hat wiederholt darauf hingewiesen, dass sich

derartige unterrichtliche Konfigurationen motivational günstig und daher störungspräventiv auswirken können (Hillenbrand, 2011; Lipowsky, 2009; Textor, 2009).

Allerdings wird auch immer wieder davor gewarnt, dass Schüler/innen mit ungünstigen Lern- und Verhaltensdispositionen bei ungenügender Strukturierung durch die Lehrperson in solchen Settings überfordert sind und in der Folge vermehrt deviantes Verhalten zeigen. Deshalb bedürfen auch geöffnete Settings einer angemessenen Strukturierung und Lernbegleitung (Brügelmann, 1997; Jang et al., 2010; Jaumann-Graumann, 2000). Diese Gratwanderung zwischen Strukturierung und Autonomieförderung stellt keine triviale Aufgabe dar für die Lehrpersonen.

Verallgemeinernd lässt sich festhalten, dass die von der Forschung identifizierten *Qualitätsdimensionen guten Unterrichts* (Helmke, 2009; Reusser, 2011) eine zentrale Orientierungshilfe darstellen bei der Suche nach einem gelingenden Umgang mit unterrichtlicher Devianz (Textor, 2007, 2009). Wettstein et al. (2010) resümieren treffend:

Die Beobachtungen deuten auf eine sich gegenseitig verstärkende Wechselwirkung zwischen mangelnder didaktischer Unterrichtsplanung und zunehmenden Unterrichtsstörungen (lange Wechselphasen, off-task-Verhalten der Schüler, aggressives und oppositionelles Schülerverhalten) hin. *Erfolgreiche Lehrkräfte* steuern den Unterricht in erster Linie über anregende und dem individuellen Lernniveau angepasste Lerninhalte. Sie erkennen problematisches Schülerverhalten früh und reagieren präventiv, bevor sich Störungen ausweiten. In seltenen Fällen greifen sie kurz und effizient auf disziplinierende Massnahmen zurück, ohne jedoch die Situation zu dramatisieren. Andere Lehrkräfte verstricken sich in lange Aushandlungen, moralisieren, klagen an und schaffen damit ungewollt einen Nährboden für einen gestörten Unterricht. (S. 103 f.)

Ergänzend und mit Blick auf *Angebots-Nutzungs-Modelle* (Reusser & Pauli, 2010) ist zu erwähnen, dass die Lernqualität unterrichtlicher Interaktionen immer auch vom Lern- und Sozialverhalten der Schüler/innen in einer Klasse abhängig ist. Deshalb kann es auch bei einem

qualitativ hochwertigen Angebots einer Lehrperson vorkommen, dass sich Kinder dennoch deviant verhalten.

Contagion-Effekte können hierbei eine Rolle spielen, also dass sich Schüler/innen von deviantem Verhalten ihrer Mitschüler/innen anstecken lassen (Müller, 2010, 2011, 2013; Müller et al., 2013; Müller, Hofmann & Studer, 2012). Denn in jeder Schulklasse befinden sich fast immer auch potenzielle Komplizen, Mitläufer oder Schaulustige, die sich über bestimmte Formen unterrichtlicher Devianz amüsieren (Breidenstein, 2006).

2.2.4. Unterrichtliche Devianz als Folge von Etikettierung?

Mit den in Kap. 2.2.2 skizzierten unterrichtlichen Verhaltenserwartungen wurde eine *kriteriale Bezugsnorm* definiert, wonach Schülerverhalten aus einer unbeteiligten, neutralen Perspektive als normkonform oder deviant beurteilt werden kann. Jedoch nehmen die am Unterrichtsgeschehen beteiligten Lehrpersonen und Schüler/innen das Verhalten auf der Basis ihrer subjektiven, voreingenommenen Perspektiven wahr – häufig dabei unbewusst und affektgesteuert (Fishbein & Ajzen, 2010; Hofer, 1986). Ob sie es als deviant beurteilen, hängt mithin auch von ihren subjektiven (normativen) Vorstellungen und Valenzen ab, die wiederum vom erweiterten sozialen Umfeld beeinflusst werden (Lamnek, 2007), z.B. von Erwartungen der Schulleitung oder der Eltern, oder von den Peers.

Demzufolge lässt sich kaum objektiv beurteilen, welche Verhaltensweisen normkonform und welche deviant sind, weil die Beurteilung immer (auch) subjektiv und situationsabhängig ist. Diese Schlussfolgerung entspricht einem zentralen Argument des *Labeling Approachs*⁹ (Becker, 1963/1981; Hargreaves, Hester & Mellor, 1981; Homfeldt, 1974).

⁹ Etikettierungstheorie und Labeling Approach werden in dieser Arbeit synonym verwendet.

Radikale Vertreter/innen dieses Ansatzes interessieren sich daher gar nicht für gezeigtes Verhalten, sondern für die asymmetrische Verteilung der Definitionsmacht unter den Interaktionsteilnehmenden sowie für deren *Motive bei der Zuschreibung von Devianz* (Stallberg & Springer, 1983). Dabei wird unterstellt, dass machtvoll soziale Akteure wie Lehrpersonen Normabweichung verursachen, "indem sie Regeln aufstellen, deren Übertretung Devianz bedeutet, und indem sie diese Regeln auf bestimmte Menschen anwenden und sie zu Outsidern abstempeln. Aus dieser Sicht ist Devianz nicht ein Merkmal der Tat, die jemand begeht, sondern die Konsequenz der Anwendung von Regeln und Sanktionen durch die anderen" (Smale, 1980, S. 35 f.).

Gemäss dem Labeling Approach könnte beispielsweise erforscht werden, ob sich Lehrpersonen aufgrund einer "Sehnsucht nach der homogenen Lerngruppe" (Tillmann, 2008, S. 38) bei der Zuschreibung von Verhaltensauffälligkeit vom Wunsch leiten lassen, dass störende Schüler/innen *wieder vom Regelschulbetrieb separiert* werden, wie es vor der bildungspolitisch angestossenen Reform zu vermehrter Integration üblich war. Solche Wunschvorstellungen wurden in der Online-Befragung des Heterogenitäts-Projekts (Reusser et al., 2013) tatsächlich von einzelnen Lehrpersonen angemerkt. Beispielsweise verlangt eine Speziallehrkraft der Mittelstufe "unkomplizierte und unbürokratische Wege, um verhaltensauffällige Schüler und Schülerinnen in eine Sonderschule zu schicken, wenn verschiedene Fachleute zum selben Ergebnis kommen" (Pappa, 2012, S. 67).

Diese *radikale Auslegung* des Labeling Approachs zeichnet sich im Vergleich zu einer rein individualpsychologischen Suchlinie im Wesentlichen dadurch aus, dass das personalisierende, latent schuldzuweisende Problemverständnis von der Seite der Störer auf die Seite der Gestörten verschoben wird: Es interessieren nicht mehr individuelle Eigenschaften der

Schüler/innen (z.B. ADHS), sondern personale Merkmale der etikettierenden Lehrpersonen (Ulich, 1980). Beide Ansätze sind in ihrer Reinform *reduktionistisch*, weil sie jeweils nur eine Seite des Problemzusammenhangs fokussieren, welcher dadurch nur unvollständig erklärt wird.

Daher wird im Folgenden von einer *pragmatischen Variante* des Labeling Approachs (Dellwing, 2011) ausgegangen: Anhand einer kriterialen Bezugsnorm (Kap. 2.2.2) wird aus der Aussenperspektive festgelegt, welche unterrichtlichen Verhaltensweisen deviant sind. Es interessiert, welche Schüler/innen sich gemäss dieser Festlegung wie oft normabweichend verhalten – und als wie stark störend Lehrpersonen und Schüler/innen diese unterrichtliche Devianz empfinden (Kap. 2.2.5).

2.2.5. Subjektives Störungsempfinden: Wahrnehmungspsychologische Ansätze

Höhn (1967/1980) konnte nachweisen, dass viele Lehrpersonen bei "schlechten Schülern" neben ungenügenden Fachleistungen eine *ganze Reihe ungünstiger Eigenschaften* wahrnehmen: Unaufmerksamkeit, mangelndes Interesse an der Schule, Langsamkeit, Unordentlichkeit, moralische Minderwertigkeit. Demzufolge scheinen manche Lehrpersonen geradezu zu erwarten, dass sich schlechte Schüler/innen deviant verhalten.

Immer wieder lässt sich ein deutlicher Halo-Effekt feststellen, der schlechte Schüler wird nur noch negativ gesehen, manchmal bis in Äußerlichkeiten hinein. Am meisten ist dies da der Fall, wo er sich dem Lehrer gegenüber ungezogen und widersetzlich verhält. Der brave, schüchterne und ängstliche schlechte Schüler wird viel milder beurteilt. (Höhn, 1967/1980, S. 105)

Mit ähnlichem Erkenntnisinteresse untersuchte Hofer (1986) *Wahrnehmungsleistungen* von Lehrpersonen. Er argumentiert, dass das menschliche Auffassungsvermögen in Schulklassen mit rund 20 Kindern von der Gesamtheit aller sozialer Prozesse naturgemäss überfordert ist. Daher nehmen Lehrpersonen unbewusst eine Komplexitätsreduktion vor, indem sie alle

Schüler/innen ihrer Klasse einer geringeren Anzahl mentaler Prototypen zuordnen (z.B. Klassenprimus, Störer). Diese *Kategorisierung der Schüler/innen* weist aber gewisse Tücken auf, denn sie erfolgt oftmals auf der Basis erster, oberflächlicher Eindrücke am Anfang des Schuljahres und bleibt hernach weitgehend stabil (Hofer, 1986).

Weil die Situations- und Personenwahrnehmung von mentalen Skripts, Erwartungen und Vorurteilen geprägt ist, können diese (mehr oder weniger zutreffenden) Schülerkategorien dazu führen, dass die Lehrpersonen vermehrt diejenigen Eigenschaften der Kinder wahrnehmen, welche ihren Kategorien entsprechen. Andere Eigenschaften werden zunehmend ignoriert (Hofer, 1986). So kann die Wahrnehmung von "Störern" durch die Lehrpersonen aufgrund von Vorverurteilungen oder Ängsten zu einem *realitätsfernen Eindruck* verzerrt werden, welcher eine vergleichsweise unnachsichtige, strenge Behandlung nach sich zieht (Hofer, 1986; Holtappels, 2000). Aus solchen dysfunktionalen, gestörten Interaktionsmustern kann eine höchst konfliktreiche Beziehung resultieren, welche die Schüler/innen noch Jahre später in negativer Erinnerung haben (Krumm, 2003; Krumm & Weiß, 2006).

Makarova, Herzog und Schönbächler (2014) untersuchten das allgemeine Störungsniveau in einzelnen Klassen als *Kontextbedingung des Störungsempfindens* von Lehrpersonen, Schülerinnen und Schülern. Es zeigte sich, dass vereinzelte undisziplinierte Verhaltensweisen in "Vielstörerklassen" (häufige Disziplinprobleme) als weniger problematisch erachtet werden als in "Wenigstörerklassen" (seltene Disziplinprobleme). Dieser Unterschied lässt sich vermutlich darauf zurückführen, dass sich die Lehrpersonen und Schüler/innen der Vielstörerklassen mit der Zeit an ein gewisses Problemausmass (z.B. Lärmpegel) gewöhnt haben, sodass einzelne Vorfälle in ihrer Wahrnehmung nicht als störende Besonderheit herausragen.

Ähnliche *Referenzgruppeneffekte* sind aus der Selbstkonzeptforschung bekannt. Mit Blick auf den empirisch gut abgesicherten Big-Fish-Little-Pond-Effekt (Möller & Trautwein, 2009) ist anzunehmen, dass sich Lernende in Vielstörerklassen als "brave" Schüler/innen einschätzen, obwohl sie sich zuweilen deviant verhalten. Dasselbe gilt vermutlich für die Aussenperspektive, sodass mit umgekehrten Vorzeichen geschlussfolgert werden kann, dass Lernende in Wenigstörerklassen schon bei vereinzelt bagatelhaften Normabweichungen als "Störenfriede" wahrgenommen werden, während sie in einer Vielstörerklasse weniger negativ auffallen würden.

Nach theoretischen Überlegungen dürften auch *Merkmale des Unterrichts* das Störungsempfinden der Lehrpersonen und der Schüler/innen beeinflussen, etwa die *didaktische Inszenierung* (Reusser, 2008): Es ist anzunehmen, dass Lärm oder motorische Unruhe im Klassenunterricht oder während Phasen der stillen Einzelarbeit mehr stören als in Inszenierungsformen, in welchen ein Grundpegel an kommunikativer und motorischer Aktivität dazugehört – etwa im Rahmen der Erweiterten Lehr-Lernformen (Crocì et al., 1995).

In diesem Zusammenhang hat Zevenbergen (2001) anhand einer ethnographischen Studie herausgearbeitet, dass die im *Klassengespräch* praktizierte triadische Gesprächskultur (Lehrperson → Schüler/in A → Lehrperson → Schüler/in B etc.) vergleichsweise *rigide Gesprächsregeln* voraussetzt, sodass bestimmte Verhaltensweisen wie etwa Wortmeldungen ohne vorgängiges Aufstrecken als störend wahrgenommen werden, während diese spezifischen Regeln in anderen unterrichtlichen Settings (z.B. Gruppenarbeit) gar nicht gelten. Nach meiner Recherche ist dieser Aspekt bislang aber nur sehr spärlich empirisch untersucht worden.

Besser erforscht ist, wie *Personenmerkmale der Wahrnehmenden* ihr Störungsempfinden beeinflussen. So empfinden es zwar die meisten Lehrpersonen als belastend, wenn sich ihre Schüler/innen undiszipliniert verhalten (Kokkinos, 2007; Krause et al., 2011; Little, 2005), doch die von einer Lehrperson subjektiv empfundene Störungsintensität hängt auch mit individualpsychologischen Merkmalen zusammen (Hamre, Pianta, Downer & Mashburn, 2008; Sherman, Rasmussen & Baydala, 2008), etwa mit ihren *Selbstwirksamkeitsüberzeugungen*: abnehmende Selbstwirksamkeit geht mit zunehmendem Störungsempfinden einher (Arbuckle & Little, 2004; Dicke et al., 2014; Grosse Siestrup, 2008).

Gemäss bisheriger Forschung empfinden auch viele *Schüler/innen* deviantes Verhalten ihrer Mitschüler/innen als störend (Infantino & Little, 2005), wenngleich sie im Durchschnitt weniger empfindlich als die Lehrpersonen reagieren (Montuoro & Lewis, 2015; Pfitzner & Schoppek, 2000). Diese Differenz lässt sich vermutlich zu einem Teil mit *rollenspezifischen Voraussetzungen* der Störungswahrnehmung erklären (Wettstein et al., 2018): Lehrpersonen sind von Berufs wegen dem Wohlergehen und dem Lernerfolg aller Schüler/innen ihrer Klasse verpflichtet – ihre pädagogischen und didaktischen Massnahmen sollten primär diesen Zielen dienen. Demgegenüber verantworten die Schüler/innen hauptsächlich ihr eigenes (Lern-)Verhalten, wofür sie eine motivationale Grundlage benötigen, wozu auch ein Minimum an Unterhaltung im Unterricht gehören kann (Breidenstein, 2006). Diese rollenspezifischen Voraussetzungen gehen mit unterschiedlichen *Valenzen, Bedürfnissen, Überzeugungen und normativen Erwartungen* einher, welche die Wahrnehmung, Deutung und Bewertung unterrichtlicher Devianz beeinflussen (Scherzinger, Wettstein & Wyler, 2017).

Ausserdem bedingen *differenzielle Opportunitäten die Wahrnehmbarkeit* von Devianz. Beispielsweise gehört die Früherkennung kleinerer Undiszipliniertheiten zur Klassenführung, damit Lehrpersonen niederschwellig intervenieren können. Die Schüler/innen hingegen richten ihre Aufmerksamkeit weniger auf solche Bagatellen, sodass sie wahrscheinlich gar nicht alle bemerken – insbesondere, wenn sie sich weit abseits ihres Sitzplatzes ereignen. Umgekehrt dürfte das stärker geächtete dissoziale Verhalten eher verdeckt gezeigt werden, z.B. während Gruppenarbeitsphasen ausserhalb des Klassenzimmers. In solchen Fällen können es nur wenige Mitschüler/innen mitbekommen, nicht aber die Lehrperson. Dies kann zur Folge haben, dass sich die Störungswahrnehmung der Lehrperson und der Schüler/innen innerhalb von Schulklassen stark unterscheidet (Wettstein, 2013; Wettstein, Ramseier, Scherzinger & Gasser, 2016).

Ähnliche Divergenzen zwischen den *verschiedenen Perspektiven* konstatiert auch Preuss-Lausitz (2005), der Schüler/innen mit besonderem Förderbedarf im Bereich der emotionalen und sozialen Entwicklung (emsoz) untersuchte:

Festzuhalten bleibt, dass in der Selbstwahrnehmung der emsoz-Schüler Probleme mit Gleichaltrigen, emotionale Probleme und Verhaltensauffälligkeiten im Vordergrund stehen. Das eigene prosoziale Verhalten wird als relativ unauffällig erlebt. Diese Einschätzung steht im Gegensatz zur Beurteilung durch die Lehrer. Die Einschätzung der Lehrer [...] wird vor allem durch die Wahrnehmung von Verhaltensauffälligkeiten und Hyperaktivität bestimmt. Lehrer sehen außerdem mangelndes prosoziales Verhalten als Problem. (S. 133)

In Bezug auf das *subjektive Störungsempfinden der Schüler/innen* hat das Forschungsprojekt "Klassenmanagement und kulturelle Heterogenität" (Makarova, Schönbächler & Herzog, 2008) interessante Ergebnisse erbracht: Nach der standardisierten Befragung der Gesamtstichprobe (N = 225 Klassen, 4'394 Schüler/innen) wurden in einer qualitativen Vertiefungsstudie 176 ausgewählte Schüler/innen interviewt, wie sie unterrichtliche Devianz

erleben (Schönbächler et al., 2009): Einige Schüler/innen beklagten sich über ein *kognitives* Störungsempfinden, etwa dass sie aufgrund von Disziplinproblemen ihrer Mitschüler/innen abgelenkt werden (Schönbächler et al., 2009). Bei manchen Schülerinnen und Schülern löst dies zudem ein *affektives* Störungsempfinden aus – sie ärgern sich beispielsweise über den damit einhergehenden Zeitverlust (Schönbächler et al., 2009). Schwerwiegende Belastungen, Frustration, Wut, Angst oder Niedergeschlagenheit werden dagegen kaum beklagt.

Verschiedene Studien zeigen, dass die Schüler/innen *interindividuell verschieden* auf unterrichtliche Devianz ihrer Mitschüler/innen reagieren. Einige fühlen sich überhaupt nicht gestört, manche erleben Devianz sogar als unterhaltsam (Breidenstein, 2006; Montuoro & Lewis, 2015; Schönbächler et al., 2009). Alles in allem bewegt sich das durchschnittliche Störungsempfinden der Schüler/innen *tendenziell auf tiefem Niveau*. Dieser Befund lässt sich wahrscheinlich damit erklären, dass gravierende Formen unterrichtlicher Devianz, welche gegen andere Schüler/innen gerichtet ist (z.B. physische Gewalt, Mobbing), nachweislich sehr selten gezeigt werden (Beaman et al., 2007; Berg et al., 1998; Crawshaw, 2015; Harrison et al., 2012; Kyriacou et al., 2007; Müller et al., 2013; Schönbächler et al., 2009).

Darüber hinaus ist anzunehmen, dass Lehrpersonen und Schüler/innen Unterschiede in der Auftretenshäufigkeit verschiedener Devianzformen berücksichtigen und gewichten, wenn sie den Störungsgrad derselben beurteilen sollen. Das würde den kontraintuitiven Befund erklären, dass sie vergleichsweise bagatellhafte *Disziplinprobleme oftmals als störendste Form* unterrichtlicher Devianz beurteilen – nicht etwa dissoziales, gewalttätiges Verhalten (Beaman et al., 2007).

Allerdings fallen bei näherer Betrachtung der zitierten Studien einige Inkonsistenzen auf, welche Crawshaw (2015) in seinem Review auf die heterogene Methodologie zurückführt. Nach Hoyt und Kerns (1999) lassen sich solche Inkonsistenzen auch mit *Rater-Effekten* erklären, welche die Wahrnehmung, Deutung und Beurteilung des gezeigten Verhalten beeinflussen. Dazu gehören beispielsweise die bereits erwähnten Halo-Effekte (Höhn, 1967/1980), Referenzgruppeneffekte (Makarova et al., 2014), an Vorurteilen orientierte, selektive Aufmerksamkeit (Hofer, 1986), rollenspezifische Wahrnehmungsleistungen (Scherzinger et al., 2017; Wettstein et al., 2018), durch Labels wie ADHS beeinflusste Wahrnehmung der Schüler/innen (Ohan, Visser, Strain & Allen, 2011) oder selbstwertdienliches Antwortverhalten von Lehrpersonen, Schülerinnen und Schülern (Bagatellisieren vs. Dramatisieren) bei Umfragen (Fishbein & Ajzen, 2010).

Ausgehend von dieser facettenreichen, nicht durchweg konsistenten Befundlage wird das Störungsempfinden der Lehrpersonen und Schüler/innen bei unterrichtlicher Devianz im SUGUS-Projekt im Sinne einer *subjektiven Erlebensqualität* verstanden. Die primären Bedingungen des Störungsempfindens stellen die Auftretenshäufigkeit und der Schweregrad devianten Schülerverhaltens dar. Darüber hinaus dürften aber auch Personenmerkmale der Wahrnehmenden sowie Merkmale des Unterrichts die Intensität und die Qualität des Störungsempfindens beeinflussen.

Bei meiner Recherche habe ich keine erziehungswissenschaftliche Quelle gefunden, welche schwerpunktmässig theoretisch klar unterscheidbare Qualitäten der Rezeption von Unterrichtsstörungen thematisiert. Vereinzelte Forschungsarbeiten (z.B. Schönbächler et al.,

2009) bringen jedoch implizit zum Ausdruck, dass das Störungsempfinden folgende *zwei Hauptkategorien* umfasst:

(i) Kognitives Störungsempfinden

z.B. "Ich werde abgelenkt", "Meine Konzentration wird beeinträchtigt"

(ii) Affektives Störungsempfinden

z.B. "Ich bin genervt", "Ich bin verärgert", "Ich bin wütend"

Die Gliederung des Störungsempfindens in diese zwei Kategorien lässt sich mit Argumenten aus der *Theorie und Forschung zu Einstellungen* begründen. Fishbein und Ajzen (2010) definieren Einstellungen als latente Dispositionen, aufgrund derer Individuen auf bestimmte psychologische Objekte in einer bestimmten Tendenz reagieren, z.B. mit Unbehagen oder Freude. Diese psychologischen Objekte können alle möglichen Aspekte der erfahrbaren Welt dieser Individuen darstellen, z.B. Verhalten. Manche Verhaltensweisen erachtet man als günstig oder angenehm, weshalb man auch selber dazu neigt, solches Verhalten zu zeigen. Anderen Verhaltensweisen steht man skeptisch gegenüber, sodass man versucht, diese zu vermeiden. Wird man durch andere Individuen mit solchem Verhalten konfrontiert, löst dies – entsprechend der skeptischen Einstellung – tendenziell negative Reaktionen auf unterschiedlichen Ebenen aus.

Rajecki (1990) unterscheidet drei solcher Ebenen, wovon sich die ersten zwei mit den zuvor genannten Dimensionen des Störungsempfindens sinngemäss decken: (1) Die *kognitive Ebene* konstituiert die informationale Deutung. Die meisten Schüler/innen realisieren, dass eine Unterrichtsstörung ihre Konzentration beeinträchtigt und sie vom Lernen abhält (Schönbächler et al., 2009). (2) Die *affektive Ebene* konstituiert die emotionale Bewertung. Schüler/innen, die

ungestört im Unterricht arbeiten wollen, empfinden es als ärgerlich, wenn aufgrund devianten Verhaltens ihrer Mitschüler/innen Zeit verloren geht (Schönbächler et al., 2009). (3) Die *behaviorale Ebene* konstituiert sodann verhaltensbezogene Reaktionstendenzen. Es gibt Lehrpersonen, die sehr aufbrausend auf Störungen reagieren, andere bleiben dagegen sehr gelassen (Makarova et al., 2014).

Zwischenfazit: Viele Lehrpersonen empfinden unterrichtliche Devianz als belastend, in Extremfällen entwickelt sich daraus sogar ein Gesundheitsrisiko (Burnout). Auch viele Schüler/innen empfinden deviantes Verhalten ihrer Mitschüler/innen als störend. Doch die subjektiv empfundene Intensität und Qualität (kognitiv vs. affektiv) einer Unterrichtsstörung wird nicht monokausal von deviantem Schülerverhalten determiniert. Auch Kontextbedingungen sowie Personenmerkmale der Wahrnehmenden haben einen Einfluss. Mit anderen Worten: *Nicht alle am Unterricht Beteiligten empfinden deviantes Schülerverhalten als gleichermassen störend.*

2.2.6. Theorieintegration: Interaktionistisches Modell gestörten Unterrichts

Unterrichtsstörungen werden häufig unter dem Aspekt der *Produktion* als primär schülerseitiges Problem erlebt und konzeptualisiert, was die Skepsis gegenüber der Integration von Schülerinnen und Schülern mit Verhaltensauffälligkeiten (Kap. 2.1.4) zu einem wesentlichen Teil erklären dürfte. Ein anderes Problemverständnis rückt demgegenüber die differenzielle *Rezeption gestörten Unterrichts* von Lehrpersonen, Schülerinnen und Schülern in den Vordergrund. Im SUGUS Projekt wird der Versuch unternommen, beide Zugänge in der theoretischen Konzeptualisierung sowie im methodischen Design zu integrieren. Vor diesem Hintergrund wurde im Rahmen von Beitrag II der Dissertation (Eckstein, Grob, et al., 2016) in Anlehnung an Wettstein (2012) und Nickel (1985) das *interaktionistische Theoriemodell* der SUGUS Studie (Abbildung 2) entwickelt. Das Modell lässt sich mit folgenden drei Kernaussagen zusammenfassen:

- i. *Unterrichtliche Devianz*: Gemessen an einer kriterialen Bezugsnorm stellt bestimmtes Schülerverhalten in bestimmten unterrichtlichen Situationen eine Normabweichung dar (Disziplinprobleme, Dissozialität).
- ii. *Subjektives Störungsempfinden*: Unterrichtliche Devianz wird von Lehrpersonen, Schülerinnen und Schülern tendenziell als störend empfunden, wobei Intensität und Qualität (kognitiv, affektiv) des Störungsempfindens interindividuell variieren.
- iii. *Multideterminiertheit des Problemzusammenhangs*: Personale Merkmale von Lehrpersonen, Schülerinnen und Schülern sowie Kontextmerkmale des Unterrichts bedingen sowohl die Auftretenshäufigkeit devianten Verhaltens als auch die Intensität und die Qualität des subjektiven Störungsempfindens.

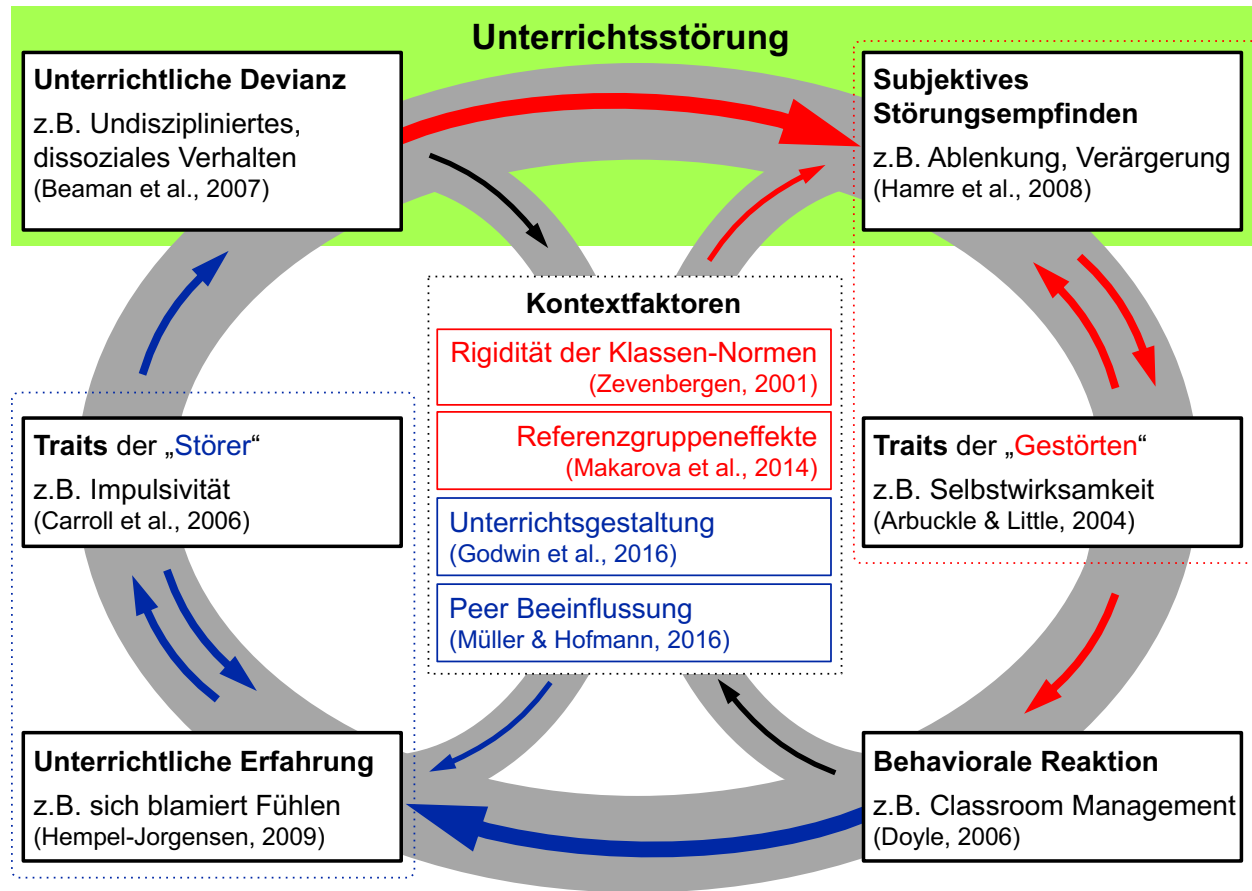


Abbildung 2. Interaktionistisches Theoriemodell der Produktion und Rezeption gestörten Unterrichts.

Die *zirkuläre Struktur* des Modells veranschaulicht eine "Störung des Regelkreises der Person-Umwelt-Beziehung" (Seitz & Stein, 2010, S. 920), wonach sich die Produktion und die Rezeption gestörten Unterrichts im zeitlichen Verlauf wechselwirksam beeinflussen: Aktuelles Schülerverhalten sowie dessen Wahrnehmung, Deutung und Bewertung durch Lehrpersonen und Mitschüler/innen werden von früheren Interaktionen beeinflusst (Rosemann & Bielski, 2001; Schweer & Thies, 2000). Weitere detaillierte Ausführungen zum Modell finden sich in der Originalarbeit (Eckstein, Grob, et al., 2016) sowie in einer geringfügig überarbeiteten Fassung in Beitrag III der Dissertation (Eckstein, 2018).

2.2.7. Konzeptualisierung des Forschungsdesigns der quantitativen SUGUS-Studie

Ausgehend vom dargelegten interaktionistischen Problemverständnis sollte die methodische Konzeptualisierung der quantitativen SUGUS-Studie zwei Zielen dienen: Das Forschungsdesign sollte es einerseits erlauben, Formen und Grade devianten *Schülerverhaltens möglichst objektiv* zu bestimmen, andererseits sollte das *subjektive Störungsempfinden* der beteiligten Lehrpersonen und Schüler/innen aus ihrer jeweiligen Perspektive erfasst werden.

Als *Zielpopulation* wurde die Mittelstufe der Primarschule bestimmt. Vor dem Hintergrund der referierten Forschung wurde angenommen, dass einige Schüler/innen der Mittelstufe zwar etwas schwerwiegendere Formen unterrichtlicher Devianz zeigen als auf der Unterstufe, aber noch nicht die gravierendsten Formen, welche manchmal auf der Oberstufe vorkommen (z.B. Delinquenz, Drogenkonsum). Damit einhergehend wurde zudem vermutet, dass auch in Bezug auf das subjektive Störungsempfinden auf der Mittelstufe viel Varianz zu finden sei.

Weiter wurde festgelegt, eine für die Nordostschweiz repräsentative Stichprobe mittels statistischer Verfahren mehrebenenanalytisch zu untersuchen. Entsprechend haben wir uns beim Studiendesign an einer *Zielgrösse von ca. 100 Schulklassen der Mittelstufe* für den Stichprobenumfang orientiert. Mit Blick auf die forschungsökonomischen Möglichkeiten und Einschränkungen wurde abgewogen, welches *Datenerhebungsverfahren* sich am besten eignet, um diese Ziele zu erreichen. Als zweckdienlichste Lösung wurde eine *multiperspektivisch angelegte, standardisierte Befragung*¹⁰ erachtet und in einer Pilotstudie erprobt.

¹⁰ Auch in Erwägung gezogen, letztlich aber nicht eingesetzt wurde eine videogestützte Beobachtung. Dafür sprach, dass unbeteiligte Beobachter/innen das Schülerverhalten objektiver einschätzen können als beteiligte Lehrpersonen und Schüler/innen. Dagegen sprach jedoch, dass mit den videographierten Lektionen lediglich von aussen sichtbare "Momentaufnahmen" (Wettstein et al., 2018,

Es wurden ganze Schulklassen rekrutiert. Befragt wurden die Klassenlehrperson und alle Schüler/innen der Klasse mit elterlicher Teilnahmeerlaubnis. Die *Datenerhebung fand zu zwei Zeitpunkten* (t_1 , t_2) jeweils während einer Lektion im Abstand von einer Woche unter der Aufsicht von Projektmitarbeitenden statt. Die Teilnehmenden erhielten einen *personalisierten Fragebogen*, welcher auf dem Deckblatt mit einer namentlich beschrifteten, ablösbaren Etikette versehen war. Der Fragebogen enthielt einen allgemeinen Teil zum Unterricht sowie einen besonderen Teil zu einzelnen *Zielschülerinnen und Zielschülern*.

Die Zielschüler/innen wurden aus drei Perspektiven mit bedeutungsgleichen Indikatoren beschrieben: von ihrer Klassenlehrperson (Teacher-Ratings); von mehreren, zufällig ausgewählten Mitschülerinnen oder Mitschülern (Peer-Ratings); von sich selber (Self-Ratings). Damit die Befragten wussten, wer in welchem Abschnitt zu beschreiben war, wurden wiederum namentlich beschriftete Etiketten verwendet. Zur Anonymisierung wurden sämtliche Etiketten am Ende der Befragung entfernt und entsorgt. Um eine korrekte Zuordnung der Daten zu gewährleisten, wurden die Fragebögen mit anonymen Personen-Identifikationsnummern (PIDs) für die Rater und Zielschüler/innen versehen.

Weil es sich bei der Beschreibung der Zielschüler/innen um *sensible Daten* handelt, wurden die Fragebögen an Einzelarbeitsplätzen mit Sichtschutz beantwortet. Zur Minimierung von Stigmatisierungseffekten wurden auch wohlwollend formulierte Items eingesetzt, welche nachträglich invertiert wurden. Die Projektmitarbeitenden berichteten von einem durchweg reibungslosen Ablauf und von einer guten Stimmung unter den Kindern nach der Befragung.

S. 108) Eingang in die Analysen finden. Darüber hinaus hätte das subjektive Störungsempfinden der Beteiligten mit einer anderen Methode erfasst werden müssen, etwa anhand einer ergänzenden Befragung. Ein solcher Methodenmix wurde als zu aufwändig verworfen, zumal neben der quantitativen Teilstudie bereits eine qualitative Vertiefungsstudie im SUGUS-Projekt geplant war.

2.2.8. Entwicklung und Pretesting der Instrumente

Gemäss den vorgängigen Recherchen hätte kein verfügbares Instrument bei der Befragung unverändert eingesetzt werden können. Daher wurde der Fragebogen grösstenteils selber entwickelt und einem *Zwei-Phasen-Pretesting* (Prüfer & Rexroth, 2000) unterzogen. In einem ersten Schritt wurden fünf Lehrpersonen und sieben Schüler/innen mit der Methode *"Retrospective Think Aloud"* (Prüfer & Rexroth, 2000, S. 8) befragt: Zunächst beantworteten sie den Fragebogens schriftlich gemäss dem vorgesehenen, standardisierten Ablauf, danach erklärten sie mündlich bei jedem Abschnitt, wie sie die Mantelfrage und den Prompt verstanden und aufgrund welcher konkreter Gedanken ihre Antworten bei den einzelnen Items zustande kamen. Dieses kognitive Pretesting lieferte wertvolle Hinweise auf mögliche Verständnisschwierigkeiten, wonach das Instrument fortschreitend überarbeitet wurde.

Als zweiter Schritt wurde im Sommer 2014 ein *Standard Pretest* mit 11 Klassen im Sommer 2014 durchgeführt (11 Klassenlehrpersonen; 239 Schüler/innen). Im Vorfeld der Erhebung bestimmte jede Lehrperson sechs Zielschüler/innen aus ihrer Klasse anhand folgender Kontrastierung: zwei Kinder, welche ihr mit Bezug auf das Verhalten positiv aufgefallen waren; zwei Kinder, die ihr negativ aufgefallen waren; zwei unauffällige, neutrale Kinder. Damit wurde ein hinreichendes Ausmass an Merkmalsvarianz sichergestellt und zugleich der Aufwand für den Pretest in einem vertretbaren Mass gehalten. So entstand eine nicht-repräsentative Teilstichprobe mit 66 Zielschülerinnen und Zielschülern, welche sich gemäss der Vorselektion durch die Lehrpersonen zu drei Zielschülergruppen ordnen lassen (positiv auffallend; negativ auffallend; neutral). Die Befragten gaben zu beiden Befragungszeitpunkten Auskunft über die Zielschüler/innen, d.h. es wurden gleiche Rater-Zielschüler-Paare zu t_1 und zu t_2 gebildet. Inhaltlich wurden zwei unterschiedliche zielschülerspezifische Aspekte erfasst.

Zum ersten Befragungszeitpunkt (t_1) wurde erfasst, wie häufig die Zielschüler/innen in den vergangenen zwei Wochen *normabweichendes Verhalten* zeigten. Das Instrument umfasste 36 Items pro Zielschüler/in, das Antwortformat bestand aus vier Kategorien ("Nie" bis "3 Mal und häufiger"). Mit dieser *niedrig-inferenten Operationalisierung* sollte der Interpretationsspielraum für die Befragten minimiert werden um Rater-Effekte zu vermindern.

Eine Woche später (t_2) berichteten die Rater, wie stark sie sich in letzter Zeit von denjenigen Zielschülerinnen und Zielschülern *gestört fühlten*, deren Verhalten sie zu t_1 eingeschätzt hatten. Die Rater beurteilten pro Zielschüler/in zehn Aussagen mittels vierstufigem Antwortformat ("Stimme gar nicht zu" bis "Stimme voll und ganz zu"). Der zeitliche Abstand sollte verhindern, dass die Befragten ihre Antworten zum Störungsempfinden bewusst ihren vorherigen Angaben zur Devianz angleichen, was den mutmasslichen Zusammenhang zwischen den zwei Konstrukten artifiziell erhöht hätte. Eine Übersicht über alle Items der Konstrukte "Devianz" und "Störungsempfinden" in der Pretest-Version findet sich in der Originalarbeit (Eckstein, Grob, et al., 2016), welche der folgenden *Fragestellung* nachging:

Eignen sich die konstruierten Skalen zur Erfassung von unterrichtlicher Devianz und subjektivem Störungsempfinden zur Untersuchung von Unterrichtsstörungen?

Die *Analysestrategie* sah vor, die faktorielle Struktur der theoretischen Konstrukte zu bestimmen sowie die Angaben der Lehrpersonen und Schüler/innen perspektivenvergleichend, deskriptiv auszuwerten. Darüber hinaus sollten diejenigen Items identifiziert werden, welche sich zur Skalierung der Konstrukte am besten eignen. Das heisst, ein Ziel der Vorstudie bestand darin, das Instrumentarium im Hinblick auf die geplante Hauptstudie zu redimensionieren, sodass in der gleichen Zeit der Befragung mehr Zielschüler/innen pro Informant beschrieben werden können.

2.2.9. Faktorenanalytische Untersuchung der im Pretest eingesetzten Instrumente

Zur Bestimmung der Dimensionalität der theoretischen Konstrukte wurden explorative (EFA) und konfirmatorische Faktorenanalysen (CFA) durchgeführt. Die Items der beiden Konstrukte "Devianz" und "Störungsempfinden" wurden zunächst separat und nach Ratingtyp¹¹ (Teacher-, Peer-Rating) getrennt auf der Ebene der individuellen Ratings untersucht. Danach wurden Modelle berechnet, in denen die beiden Konstrukte gleichzeitig geschätzt wurden.

Letztendlich hat sich eine dreifaktorielle Struktur als beste Lösung für beide Ratingtypen erwiesen: *Unterrichtliche Devianz*; *kognitives Störungsempfinden*; *affektives Störungsempfinden*. Das finale Modell der Lehrpersonen ($\chi^2=238.26$; $df=227$, n.s.; RMSEA=.027; CFI=.998) sowie das Schülermodell ($\chi^2=529.71$, $df=497$, n.s.; RMSEA=.017; CFI=.956) passen gut zu den Daten. Abbildung 3 illustriert das *finale Schülermodell*, bei welchem zusätzlich die hierarchische Datenstruktur mehrebenenanalytisch berücksichtigt wurde. Die Befunde des Modells der Lehrpersonen sind inhaltlich sehr nahe am Schülermodell, sodass im vorliegenden Kapitel nur Letzteres berichtet wird.

Ebene zwei (L2) ist die *Ebene der Zielschüler/innen* ($N_{L2}=66$), wo die Einschätzungen von durchschnittlich 3.6 Peers messfehlerbereinigt zu latenten Random Intercepts¹² aggregiert wurden. Bei den latenten Faktoren auf L2 handelt es sich um *Trait-Faktoren*, welche die von den allen Peer-Ratern geteilte Sichtweise auf die Zielschüler/innen in sich vereinen.

¹¹ Mit dem Begriff "Ratingtyp" bezeichne ich die unterschiedlichen *Arten der Ratings* (Teacher-, Peer- und Self-Ratings). Später werde ich auch den Begriff "Ratertyp" verwenden, welcher die *Personen* bezeichnet, welche die Ratings vorgenommen haben (Lehrperson, Mitschüler/in, Zielschüler/in).

¹² Im vorliegenden Abschnitt verwende ich eine andere Terminologie als in der Originalarbeit (Eckstein, Grob, et al., 2016), ich berichte aber keine neuen Analysen oder andere Ergebnisse. Es handelt sich um die Terminologie der Multitrait-Multimethod-(MTMM)-Verfahren wie in Beitrag IV der Dissertation (Eckstein, under review).

Auf Ebene 1 (L1) werden *Abweichungen der unique Peer-Ratings* ($N_{L1}=234$) von den gemittelten Scores modelliert. Technisch gesprochen handelt es sich hierbei um Unter- bzw. Überschätzungen durch einzelne Rater, welche auf L1 als *unique Methoden-Faktoren* latent modelliert werden. Anhand der Korrelationen zwischen den Methoden-Faktoren wird geschätzt, inwieweit die Rater eine *allgemeine* Tendenz aufweisen, die Zielschüler/innen konstruktübergreifend zu über- oder unterschätzen (Eckstein, under review; Eid, Lischetzke, Nussbeck & Trierweiler, 2003; Eid et al., 2008).

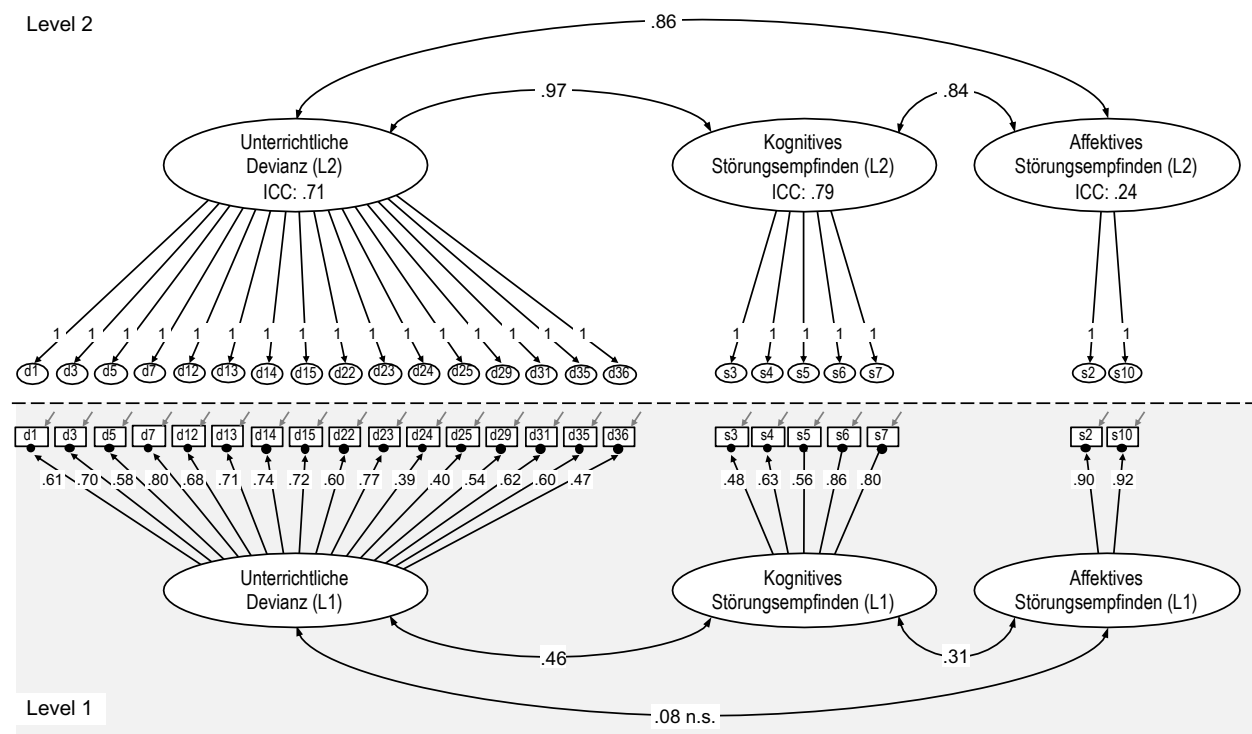


Abbildung 3. Mehrebenenmodell der Peer-Ratings (Eckstein, Grob, et al., 2016, S. 123).

Anmerkungen: Aufgrund der vereinheitlichten Darstellungsweise der SEM im vorliegenden Manteltext der Dissertation nach Muthén und Muthén (2017), weicht Abbildung 3 geringfügig ab von der Darstellung in der Originalarbeit (Eckstein, Grob, et al., 2016). Die auf L2 latent aggregierten Ratings werden als Random Intercepts ohne Messfehler dargestellt (der Fehler wurde auf L2 auf Null fixiert, weil er vollständig auf L1 ist; daher werden die Intercepts auf L2 perfekt durch die Trait-Faktoren erklärt, sodass die standardisierten Faktorladungen 1.00 betragen). Abgebildet sind die standardisierten Parameter, die alle signifikant sind mit einer maximalen Irrtumswahrscheinlichkeit von $p < .05$, ausser es ist anders vermerkt.

Die ermittelten *Intraklassenkorrelationen* (ICC 1) können als Konsistenzmass interpretiert werden¹³. Die ICC des latenten Faktors *unterrichtliche Devianz* (.71) ist sehr hoch verglichen mit den in der Unterrichtsforschung üblichen Koeffizienten um .10 – .20 (Wagner et al., 2016). Das Resultat bedeutet, dass 71 Prozent der Varianz von den Peer-Ratern geteilt wird und nur 29 Prozent der Varianz raterspezifisch ist. Anders ausgedrückt: Die einzelnen Peers haben *intersubjektiv höchst konsistent eingeschätzt*, wie häufig die Zielschüler/innen deviantes Verhalten zeigten – die Rater-Effekte sind vergleichsweise minimal. Demzufolge hat sich die niedrig-inferente Operationalisierung des Instruments gelohnt, deren Ziel darin bestand, den objektiven Kern von Unterrichtsstörungen approximativ zu bestimmen.

¹³ In Beitrag II der Dissertation (Eckstein, Grob, et al., 2016) wurden die *ICC(1)-Koeffizienten* in Anlehnung an Hox (2010) pro Konstrukt wie folgt berechnet: Trait-Faktor-Varianz (L2) geteilt durch messfehlerbereinigte Gesamtvarianz [(Trait-Faktor-Varianz (L2)) plus (unique Methoden-Faktor-Varianz (L1))]. Der resultierende ICC(1)-Koeffizient entspricht demjenigen Varianzanteil, zu welchem die multiplen Peer-Ratings pro Zielschüler/in im Mittel konsistent ausgefallen sind. Somit haben diese ICC(1)-Koeffizienten eine analoge Bedeutung wie die Konsistenz-Koeffizienten, welche in Beitrag IV der Dissertation (Eckstein, under review) in Anlehnung an Eid et al. (2008) berechnet wurden. Im Falle der messfehlerbereinigten True Scores der Peer-Ratings sind die zwei Koeffizienten vollkommen identisch, was sich mathematisch herleiten lässt.

Die *Konsistenz-Koeffizienten* der True Scores wurden pro Konstrukt wie folgt gebildet: Im Zähler wurde die Trait-Faktor-Varianz mit der quadrierten Summe der unstandardisierten Trait-Faktorladungen multipliziert (L2). Im Nenner wurde die messfehlerbereinigte Gesamtvarianz additiv gebildet: [Trait-Faktor-Varianz mal quadrierte Summe der unstandardisierten Trait-Faktorladungen (L2)] plus [unique Methoden-Faktor-Varianz mal quadrierte Summe der unstandardisierten Methoden-Faktorladungen (L1)].

Zur Vermeidung von Cluster Biases wurden die unstandardisierten Faktorladungen auf beiden Ebenen gleich gesetzt (Jak, Oort & Dolan, 2013). Das bedeutet, dass die unstandardisierten Faktorladungen der Indikatoren der Peer-Ratings auf die Trait-Faktoren (L2) und auf die Methoden-Faktoren (L1) gleiche Werte aufweisen. Deshalb können die Ladungen im Nenner ausgeklammert werden, worauf sich der Bruch um die Ladungen kürzen lässt. Nach diesem Schritt lautet die Formel zur Berechnung der Konsistenz-Koeffizienten gleich wie diejenige zur Berechnung der ICC(1): *L2-Varianz geteilt durch Gesamtvarianz (L2 + L1)*.

Auch in Bezug auf das *kognitive Störungsempfinden* sind die Peer-Ratings sehr konsistent ($ICC = .79$); nur 21 Prozent der Varianz ist raterspezifisch. Demgegenüber sind die Peer-Ratings in Bezug auf das *affektive Störungsempfinden* weit weniger konsistent – die Rater teilen nur 24 Prozent der Varianz miteinander ($ICC = .24$), der Löwenanteil der Einschätzungen (76 Prozent der Varianz) ist auf Rater-Effekte zurückzuführen. Demzufolge sind sich die Peers zwar weitgehend darin einig, inwieweit die Zielschüler/innen im Unterricht für Ablenkung sorgen, doch es ist höchst subjektiv, ob und wie stark sie sich über die Zielschüler/innen ärgern.

Die *ebenenspezifischen Interfaktorkorrelationen* des Modells sind wie folgt zu deuten: Gemäss durchschnittlicher Einschätzung der Peers (L2) werden Zielschüler/innen als zunehmend störend empfunden, je häufiger sie deviantes Verhalten zeigen – und zwar in kognitiver ($r = .97^*$) sowie in affektiver ($r = .86^*$) Hinsicht. Es wurden zwar starke Zusammenhänge erwartet, doch die Koeffizienten sind derart hoch, dass ich sie in den nachfolgenden Abschnitten noch kritisch diskutieren werde.

Auf L1 fallen die Zusammenhänge deutlich geringer aus. Demnach sind die Rater-Effekte teilweise generalisierbar, aber nicht über alle Konstrukte hinweg: Wenn einzelne Kinder einen Zielschüler hinsichtlich der Auftretenshäufigkeit unterrichtlicher Devianz überschätzen (verglichen mit den mittleren Peer-Ratings), dann erachten sie diesen Zielschüler tendenziell auch als kognitiv störender als die anderen Rater ($r = .46^*$). Beim affektiven Störungsempfinden ist das aber nicht der Fall – Über- oder Unterschätzung mit Blick auf Devianz korreliert nicht mit dem affektiven Störungsempfinden ($r = .08$).

2.2.10. Deskriptive Befunde der Vorstudie

Ausgehend von den Faktorenanalysen wurden die *Mittelwerte der drei skalierten Konstrukte* für Beitrag II der Dissertation (Eckstein, Grob, et al., 2016) manifest berechnet. Die Resultate deuten darauf hin, dass normabweichendes Verhalten in der untersuchten Stichprobe im Durchschnitt nur selten gezeigt wurde. Auch das kognitive und das affektive Störungsempfinden der befragten Lehrpersonen und Schüler/innen sind durchschnittlich gering ausgeprägt. Sämtliche über alle Zielschüler/innen gemittelten Mean Scores bewegen sich im Bereich von 0.31 – 1.19 und liegen damit deutlich unterhalb des theoretischen Mittelwerts der verwendeten Skala von 1.5. Eine *perspektivenvergleichende Darstellung* der Mittelwerte nach Rater- und Zielschülergruppe findet sich in Abbildung 4.

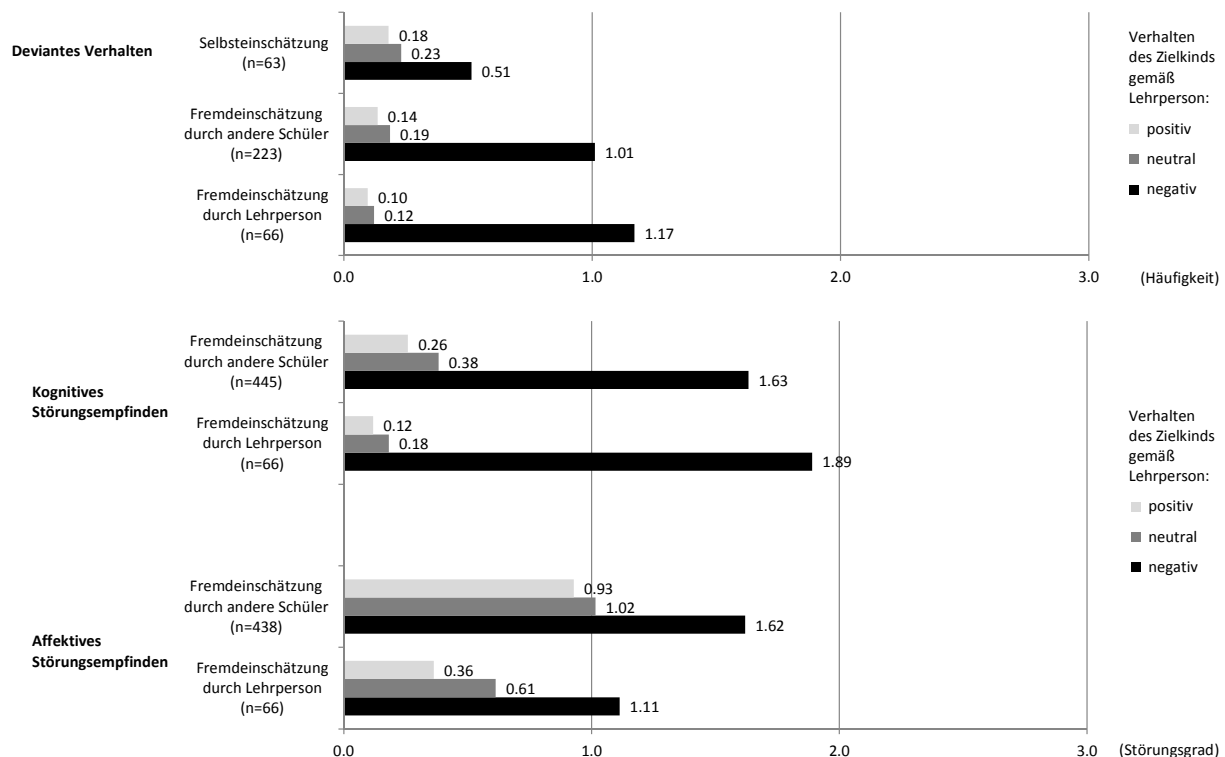


Abbildung 4. Deskriptiver Mittelwertsvergleich nach Rater- und Zielschülergruppe.

Die Darstellung zeigt, dass sich der von den Lehrpersonen konstruierte Kontrast zwischen den drei Zielschülergruppen tendenziell auch in der Perspektive der Schüler/innen spiegelt, jedoch mit leicht anderen Akzenten. Diejenigen Zielschüler/innen, welche von den Lehrpersonen vor der Befragung als "negativ auffallend" deklariert wurden, zeigen gemäss allen Ratern durchschnittlich am häufigsten deviantes Verhalten und lösen das höchste Störungsempfinden aus. Demgegenüber scheinen sich die "neutralen" und die "positiv auffallenden" Zielschüler/innen in der durchschnittlichen Wahrnehmung der drei Ratertypen weniger deutlich voneinander zu unterscheiden¹⁴.

Zusätzlich zu den Analysen für Beitrag II der Dissertation (Eckstein, Grob, et al., 2016) wurden weitere deskriptive Auswertungen auf der Ebene der einzelnen an der Studie teilnehmenden Schulklassen vorgenommen. Die Ergebnisse dieser klassenspezifischen Analysen wurden jeder teilnehmenden *Lehrperson in Form eines personalisierten Ergebnisberichts* zugestellt (Eckstein, Reusser, Grob & Hofstetter, 2015a, 2015b). Damit wurde im Rahmen des SUGUS-Projekts erstmals eine bestimmte Form der Ergebniskommunikation eingesetzt. Bei der späteren Rekrutierungsphase der SUGUS-Hauptstudie wurden den angefragten Lehrpersonen solche personalisierte Berichte als Incentive in Aussicht gestellt.

¹⁴ Die Interpretation des deskriptiven Perspektivenvergleichs im Pretest beruht auf schlussfolgernden Überlegungen. Auf eine statistische Absicherung der Mittelwertsvergleiche (z.B. mittels t-Tests, Varianzanalysen) wurde verzichtet, weil eine Generalisierung der Befunde aufgrund der nicht-repräsentativen Stichprobe nicht zulässig gewesen wäre.

Für die klassenspezifischen Analysen wurden bei den drei Konstrukten Devianz, affektives und kognitives Störungsempfinden die Mittelwerte über alle Zielschüler/innen pro Klasse je Ratingtyp berechnet. Zwei weitere Konstrukte (Reizbarkeit und Ablenkbarkeit) wurden zusätzlich berücksichtigt: "Reizbarkeit" misst, inwieweit sich Lehrpersonen und Schüler/innen von prototypischen Unterrichtsstörungen verärgern lassen; "Ablenkbarkeit" misst, wie sehr sie sich davon ablenken lassen (Eckstein, Grob & Reusser, 2015). Die Angaben der Lehrpersonen zu ihrer eigenen Reizbarkeit und Ablenkbarkeit wurden auf Individualebene analysiert, bei den Schülerinnen und Schülern wurden Mittelwerte pro Klasse berechnet.

Die Resultate dieser deskriptiven Analysen wurden in den für die Lehrpersonen erstellten Berichte in Form von Netzdiagrammen anschaulich dargestellt (Eckstein, Reusser, et al., 2015a, 2015b). Anhand der Netzdiagramme wird das wahrgenommene Problemausmass aus mehreren Perspektiven simultan dargestellt: Je grösser die abgebildete Fläche, desto gravierender die Situation in der jeweiligen Perspektive. Abbildung 5¹⁵ illustriert die Sichtweise der Lehrperson (Lp) der Klasse "Beispiel" im Vergleich mit der durchschnittlichen Perspektive ihrer Schülerinnen und Schüler (SuS).

¹⁵ Das in Abbildung 5 dargestellte Netzdiagramm beruht auf echten Daten – es ist kein fiktives Beispiel. Jedoch haben wir diesen Diagramm-Typ "Lp/SuS" den Lehrpersonen nicht zugestellt. Wir haben befürchtet, dass der darin enthaltene Perspektivenvergleich dazu verleiten könnte, Rückschlüsse auf einzelne Zielschüler/innen zu ziehen, weil diese durch die Lehrpersonen vorselektiert wurden. Stattdessen enthielten die an die Lehrpersonen abgegebenen personalisierten Ergebnisberichte (Eckstein, Reusser, et al., 2015b), nebst weiteren Angaben, ein Netzdiagramm-Typ "Lp/Lp" mit den Werten der adressierten Lehrperson im Vergleich mit den durchschnittlichen Werten aller Lehrpersonen, welche sich am Pretest beteiligt hatten.

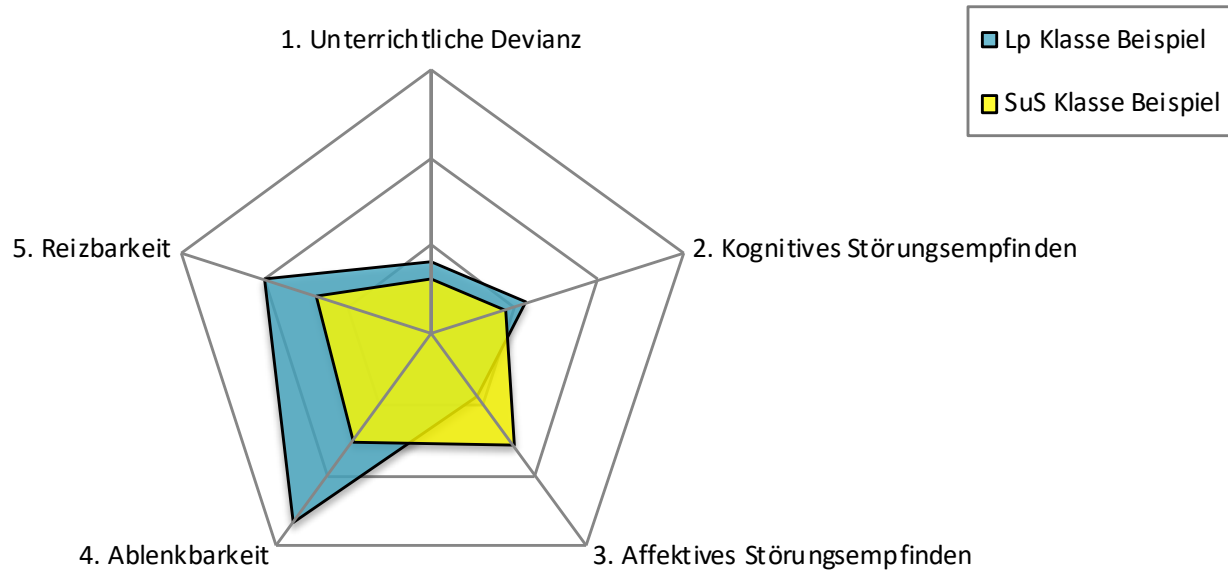


Abbildung 5. Perspektivenvergleich: Schüler/innen vs. Lehrperson innerhalb der Beispielklasse.

In diesem Netzdiagramm zeigt sich, dass die Beispiel-Lehrperson (blaue Fläche) nach eigener Auskunft in störenden Situationen generell mehr abgelenkt wird (4) und sich mehr ärgert (5) als ihre Schüler/innen (gelbe Fläche). Womöglich führt diese erhöhte Sensitivität dazu, dass die Lehrperson auch die Auftretenshäufigkeit devianten Verhaltens (1) höher eingeschätzt als ihre Schüler/innen. Darüber hinaus fühlt die sich Lehrperson von den Zielschülerinnen und Zielschülern in kognitiver Hinsicht mehr gestört als ihre Schüler/innen (2). Einzig in Bezug auf das affektive Störungsempfinden (3) gab diese Lehrperson tiefere Werte an als ihre Schüler/innen. Darin spiegelt sich eine allgemeine Tendenz, waren doch die meisten Lehrpersonen sehr zurückhaltend, negative Affekte gegenüber ihrer Schüler/innen zu berichten. Vermutlich ist diese Zurückhaltung Ausdruck eines professionellen Rollenverständnisses bzw. ihres pädagogischen Ethos. Demgegenüber äusserten sich die Schüler/innen freimütiger, wenn sie Mitschüler/innen nicht mochten oder unsympathisch fanden.

2.2.11. Zentrale Schlussfolgerungen und offen gebliebene Fragen nach der Vorstudie

Die Resultate des Pretests haben gezeigt, dass sich die theoretisch entflochtenen *Merkmalsdimensionen gestörten Unterrichts* anhand der multiperspektivisch angelegten Befragung von Lehrpersonen, Schülerinnen und Schülern gehaltvoll untersuchen lassen. Darüber hinaus konnten dank der Vorstudie wertvolle Hinweise zur Optimierung des Forschungsdesigns gesammelt werden. Die folgenden fünf Abschnitte skizzieren, welche Problemzonen identifiziert wurden und welche Fragen offen geblieben sind.

Operationalisierung. Eine erste Problemzone wurde mit Bezug auf das Konstrukt Störungsempfinden identifiziert. Es konnte nicht ausgeschlossen werden, dass gewisse Items je nach Ratingtyp Unterschiedliches messen. Mit dem Ziel paralleler Messungen musste an den Formulierungen gearbeitet werden. Eine zweite Schwachstelle wurde in der oberen Begrenzung des Antwortformats der Skala unterrichtliche Devianz vermutet ("3 Mal und häufiger"). Mit dem Ziel, auch höhere Auftretenshäufigkeiten devianten Verhaltens feiner und präziser zu erfassen, wurde auch dieses Antwortformat überarbeitet.

Sampling. Aufgrund der Vorselektion der Zielschüler/innen (positiv, negativ, neutral) durch die Lehrpersonen konnte nicht ausgeschlossen werden, dass die Lehrpersonen die Zielschüler/innen voreingenommen eingeschätzt haben. Ähnliches wurde mit Bezug auf die Peer-Ratings befürchtet. Denn es ist nicht unplausibel anzunehmen, dass die Schüler/innen mit ihrer Lehrperson durchaus darin übereinstimmen, welche Kinder in der Klasse negativ auffallen – auch wenn ihre persönliche Beziehung zu diesen Kindern neutral oder gar freundschaftlich ist. Daraus resultierte möglicherweise eine artifizielle Überzeichnung des Kontrasts der "negativ

auffallenden" Zielschülerinnen und Zielschülern gegenüber den zwei anderen Zielschülergruppen. Deshalb wurde vermutet, die hohen Korrelationen zwischen den Trait-Faktoren auf Ebene zwei im Schülermodell (Abbildung 3) seien womöglich ein Symptom dieser mutmasslichen Überzeichnung. Um solche Verzerrungen bei der SUGUS-Hauptstudie ausschliessen zu können, wurde die Samplingstrategie grundlegend überarbeitet (Kap. 2.3.2).

Dimensionalität der theoretischen Konstrukte. Es blieb vorerst unklar, ob sich die anhand der Pretest-Daten faktorenanalytisch bestimmte Dimensionalität in der Hauptstudie replizieren lassen würde. So war zum Beispiel nicht sicher, ob das Konstrukt Devianz weiterhin eindimensional skaliert werden würde, oder ob sich gemäss der theoretischen Einteilung eine zweidimensionale Skalierung realisieren liesse (Disziplinprobleme vs. Dissozialität, Kap. 2.2.5).

Deskriptives Problemausmass. Die Resultate der deskriptiven Analysen deuten auf eine durchschnittlich geringe Belastung in der Pretest-Stichprobe hin. Dieser Befund steht etwas im Widerspruch zum tendenziell alarmierenden Pressespiegel, welcher Unterrichtsstörungen im Zusammenhang mit Verhaltensauffälligkeiten als schwerwiegendes Problem der (integrativen) Schule darstellt (Kap. 2.1.4). Allerdings ist die Aussagekraft des in der Vorstudie ermittelten Resultats begrenzt, weil die Analysen auf einer nicht-repräsentativen Stichprobe beruhen. Darüber hinaus quantifizieren die Mittelwerte lediglich das durchschnittliche Störungsniveau – die Varianz wurde bislang nicht systematisch analysiert, sodass nicht ausgeschlossen werden konnte, dass die Belastung in einzelnen Klassen wesentlich höher ist, als die Mittelwerte zum Ausdruck bringen.

Schliesslich haben die elf *klassenspezifischen Analysen* nahegelegt, dass sich die Perspektiven innerhalb der einzelnen Schulklassen deutlicher unterscheiden als zwischen den Klassen. Im Pretest gelang es jedoch nicht, die Angaben aller Lehrpersonen und Schüler/innen in einem SEM simultan zur Schätzung der interessierenden Parameter heranzuziehen.

Ausgehend von den Pretest-Ergebnissen und den daraus gezogenen Schlussfolgerungen sowie den offen gebliebenen Fragen wurde im nächsten Schritt die SUGUS-Hauptstudie vorbereitet.

2.3. Beitrag III: Deskriptiver Perspektivenvergleich

*Abstract*¹⁶

Manche Schüler/innen zeigen im Unterricht gelegentlich Verhaltensweisen, die von unterrichtlichen Normen abweichen. Diese unterrichtliche Devianz wird von den beteiligten Akteuren in ihrer jeweiligen Perspektive als mehr oder weniger stark störend empfunden, woraus Unterrichtsstörungen hervorgehen können. In der Praxis wird der Aspekt der differenziellen Störungsrezeption nicht immer bewusst reflektiert. Die SUGUS-Studie untersucht Unterrichtsstörungen aus der Sichtweise von 85 Lehrpersonen und ihren 1'412 Schülerinnen und Schülern (11.7 Jahre). Sie schätzten ein, wie häufig einzelne Zielschüler/innen aus ihrer Klasse deviantes Verhalten zeigten und welches Störungsempfinden dies bei ihnen ausgelöst hat. Zudem schätzten sie ihre eigene Störungsempfindlichkeit anhand standardisierter Fallvignetten ein und gaben Auskunft über ihr allgemeines Belastungserleben bezogen auf die Schule. Die Daten wurden auf Stichprobenebene sowie für jede an der Studie teilnehmende Klasse deskriptiv perspektivenvergleichend ausgewertet. Die klassenspezifischen Resultate wurden den Lehrpersonen in Form personalisierter Berichte zurückgemeldet.

Beitrag III der Dissertation (Eckstein, 2018)¹⁷ entstand im Rahmen der SUGUS-Hauptstudie. In den Kapiteln 2.3.1 bis 2.3.3 werde ich die wichtigsten Neuerungen dieser Hauptstudie gegenüber dem zuvor besprochenen Pretest erläutern. Kapitel 2.3.4 behandelt das Datencleaning und stellt Resultate deskriptiver Basisanalysen auf Stichprobenebene vor. Anschliessend (Kap. 2.3.5) präsentiere ich Befunde klassenspezifischer Analysen, welche den Lehrpersonen zurückgemeldet wurden – das ist das Hauptthema von Beitrag III. Eine Diskussion der Grenzen dieses deskriptiven Perspektivenvergleichs schliesst das Kapitel ab.

¹⁶ Bei der Originalarbeit III handelt es sich um einen Beitrag in einem Tagungsband, dem kein Abstract vorangestellt ist. Deshalb wurde an dieser Stelle ein neues Abstract verfasst.

¹⁷ Die vollständigen bibliographischen Angaben der Originalarbeit lauten: Eckstein, B. (2018). Unterrichtsstörungen: eine Frage der Perspektive? In S. Schwab, G. Tafner, S. Luttenberger, H. Knauder & M. Reisinger (Hrsg.), *Von der Wissenschaft in die Praxis? Zum Verhältnis von Forschung und Praxis in der Bildungsforschung* (S. 78–92). Münster: Waxmann.

2.3.1. Vorbemerkungen zur SUGUS-Hauptstudie

Die in der SUGUS-Vorstudie (Kap. 2.2) gemachten Erfahrungen gaben keinen Anlass, den Theorierahmen des Projekts grundlegend zu überarbeiten. Folglich basiert auch die Hauptstudie auf dem *interaktionistischen Theoriemodell* (Abbildung 2), wonach deviantes Schülerverhalten (Produktion) und das subjektive Störungsempfinden der Lehrpersonen und Schüler/innen (Rezeption) simultan in den Blick genommen werden. Die nachfolgenden Abschnitte stützen sich somit auf die bisherigen Ausführungen zum Theorie- und Forschungsstand und werden nicht wiederholt.

Auch das *Forschungsdesign* hat sich im Rahmen des Pretests als vielversprechend erwiesen, wenn auch einzelne Elemente im Hinblick auf die Hauptstudie verbessert werden mussten. Im Anschluss an diese Vorbemerkungen werde ich im Detail darlegen, wie wir die Samplingstrategie und das Instrumentarium vor der Hauptstudie weiterentwickelt haben (Kap. 2.3.2, 2.3.3). Eine ausführliche Dokumentation der in der Hauptstudie realisierten Methoden – inklusive finaler Skalierung der Konstrukte – findet sich im technischen Bericht der quantitativen Teilstudie des SUGUS-Projekts (Eckstein et al., 2018).

Eine weitere Vorbemerkung betrifft die *zeitliche Abfolge* der für Beitrag III unternommenen Analyseschritte und damit verbundene Konsequenzen in Bezug auf die Skalierung der theoretischen Konstrukte. Wie bereits im Pretest (Eckstein, Reusser, et al., 2015a, 2015b) haben wir jeder bei der Rekrutierung angefragten Lehrperson als Anreiz und als Dank für ihre Teilnahme an der SUGUS-Hauptstudie einen personalisierten Ergebnisbericht mit Resultaten klassenspezifischer Analysen (Eckstein, Luger, Grob & Reusser, 2016a, 2016b) in Aussicht gestellt (neben einem Büchergutschein à 100 CHF). Bei der Vorstudie mussten die Lehrpersonen allerdings über ein Jahr auf ihre Berichte warten. Das lag daran, dass die

deskriptiven Analysen (Kap. 2.2.10) erst auf der Grundlage der faktorenanalytisch bestimmten Skalierung der interessierenden Konstrukte (Kap. 2.2.9) erfolgten, welche viel Zeit in Anspruch nahm. Um die an der Hauptstudie teilnehmenden Lehrpersonen nicht so lange warten zu lassen, wurde entschieden, dieses Mal die deskriptiven Analysen vorzuziehen und die Faktorenanalysen später durchzuführen. Dieser Entscheid hatte letztlich zur Folge, dass sich die in den personalisierten Ergebnisberichten untersuchten Konstrukte in ihrer Item-Zusammensetzung unterscheiden von den faktorenanalytisch bestimmten Skalen, die im späteren Beitrag IV der Dissertation verwendet wurden (vgl. Kap. 2.4.1).

2.3.2. Sampling bei der Hauptstudie

Aufgrund der mutmasslich problematischen Vorselektion der Zielschüler/innen durch die Lehrpersonen beim Pretest wurden für die Hauptstudie *alternative Samplingstrategien diskutiert*. Zunächst wurde in Betracht gezogen, etwa zehn Zielschüler/innen pro Klasse zufällig zu bestimmen. Doch mit dieser Strategie wären vermutlich einige Schüler/innen unberücksichtigt geblieben, welche von ihrer Lehrperson und/oder von ihren Peers als "Störenfriede" (Burri, 2017) wahrgenommen werden – mit Blick auf die Ziele der Studie wäre dies fatal gewesen. Zudem hätte es wohl auch die Lehrpersonen irritiert, wenn sie in der Befragung zu Unterrichtsstörungen eine Teilstichprobe von lauter "braven" Zielschülerinnen und Zielschülern hätten einschätzen müssen.

Deshalb wurde beschlossen, dass *alle Schüler/innen der an der Studie teilnehmenden Klassen zugleich Rater und Zielschüler/innen* sein sollen. Damit wurde auch die Gefahr reduziert, dass einzelne Zielschüler/innen nach der Umfrage in Gesprächen unter den Kindern besonders hervorgehoben werden würden. Die Lehrpersonen beschrieben alle Schüler/innen

ihrer Klasse (Teacher-Ratings), die Schüler/innen beschrieben sich selber (Self-Ratings) sowie vier zufällig ausgewählte Mitschüler/innen (Peer-Ratings). Die Zuteilung der Zielschüler/innen zu den Peers erfolgte randomisiert.

Nicht an der Studie teilnehmende Schüler/innen liessen wir nicht als Zielschüler/innen von ihren Peers beschreiben, weil wir vermeiden wollten, dass sie sich ausgestellt fühlen ohne selber Stellung nehmen zu können (erfahrungsgemäss haben sich viele Kinder jeweils nach der Erhebung untereinander darüber ausgetauscht, wer wen einschätzen musste). Um dennoch eine Information über die nicht-teilnehmenden Schüler/innen zu erhalten, liessen wir sie von den Lehrpersonen, mit denen wir eine vertrauliche Handhabe der sensiblen Daten vereinbarten, beschreiben.

Insgesamt haben 85 *Primarschulklassen* aus zehn Kantonen der Nordostschweiz an der quantitativen SUGUS-Hauptstudie teilgenommen. Von den 85 Klassen waren rund 90% reine 5. Klassen, knapp 10% waren jahrgangsdurchmischte Klassen der Mittelstufe. Zum Zeitpunkt der Befragung wurden die Klassen von 1'687 Schüler/innen besucht, wovon sich 1'412 Schüler/innen mit elterlicher Teilnahmeerlaubnis zur Studie angemeldet haben. Insgesamt 275 Schüler/innen haben sich von der Studie abgemeldet und wurden folglich nur von der Lehrperson beschrieben, nicht aber von ihren Mitschülerinnen und Mitschülern und auch nicht von sich selbst. Detaillierte Angaben zur Rekrutierung und zum Rücklauf sowie deskriptive Basisinformationen über die Stichprobe finden sich im technischen Bericht (Eckstein et al., 2018).

2.3.3. Weiterentwicklung des Instrumentariums nach der Vorstudie

In ihrer Grundstruktur sind die Instrumente seit der Vorstudie unverändert geblieben, insbesondere die Personalisierung der Fragebögen hat sich bewährt. (Kap. 2.2.8). Allerdings hat sich die Arbeit mit den wiederablösbaren *Etiketten als äusserst umständlich* erwiesen. Die manuelle Zuordnung der namentlich beschrifteten Etiketten mit den entsprechenden Personen-Identifikationsnummern (PIDs) gelang wohl nur deshalb fehlerfrei, weil das kleine Pretest-Sample überschaubar war. Bei der Hauptstudie waren jedoch 2'994 Fragebogen-Unikate¹⁸ zu produzieren und es mussten alleine für die Peer-Ratings zwei mal 5'648 Rater-Zielschüler/innen-Paare¹⁹ randomisiert gebildet werden. Dabei mussten die Paare immer aus der gleichen Klasse stammen und sie mussten zu t_1 und t_2 identisch sein. Darüber hinaus war zu berücksichtigen, dass die nicht-teilnehmenden Schüler/innen zwar von ihrer Lehrperson, nicht aber von ihren Peers als Zielschüler/innen eingeschätzt werden sollten. Aufgrund dieser Komplexität wurde befürchtet, dass sich bei einer manuellen Personalisierung zwangsläufig Fehler ereignen würden.

Deshalb wurden die Fragebögen bereits in der Serienproduktion *automatisiert personalisiert*: Der Seitenrand der Broschüren wurde perforiert, sodass ein Abrissstreifen entstand, worauf die namentlichen Angaben der Rater (auf dem Deckblatt) sowie der Zielschüler/innen (im Innern der Fragebögen) gedruckt wurden. Neben dem Abrissstreifen wurden die PIDs gedruckt. Nach der Befragung wurde der Streifen zur Anonymisierung entfernt und entsorgt. Zur Vorbereitung dieses Unterfangens musste eine vollständige Urliste sämtlicher Rater-Zielschüler/innen-Paare erstellt werden. Entsprechend frühzeitig musste abgeklärt werden, welche Schüler/innen mit elterlicher Erlaubnis an der Studie (nicht) teilnehmen werden.

¹⁸ 1'412 Broschüren für Schüler/innen plus 85 Broschüren für Lehrpersonen mal 2 (t_1 , t_2)

¹⁹ 1'412 Peer-Rater mal 4 Zielschüler/innen

Neben der formalen Neugestaltung wurden die Instrumente auch inhaltlich überarbeitet. Ein erstes wichtiges Massnahmenpaket betraf das Konstrukt *unterrichtliche Devianz*. Zunächst wurde eine andere Zeitform gewählt, um das vergangene Verhalten der Zielschüler/innen einschätzen zu lassen (Perfekt statt Präsens). Sodann wurden folgende *zwei Items nachträglich ausgeschlossen bzw. durch ein neues Item ersetzt*: "Verweigert die Zusammenarbeit mit einem anderen Kind" (d24); "Redet schlecht über ein anderes Kind hinter seinem Rücken" (d25). Die Items sollten dissozial-ausgrenzendes Verhalten messen. Dass dies nicht optimal gelang, deuten die vergleichsweise tiefen Faktorladungen im oben berichteten Schülermodell an (Abbildung 3). Das könnte daran liegen, dass das interessierende Verhalten in dieser spezifischen Form für viele Rater kaum sichtbar war. Das neu gebildete Ersatz-Item soll das dissozial-ausgrenzende Verhalten unabhängig davon messen, ob es offen oder verdeckt gezeigt wird: "Hat im Unterricht über ein anderes Kind gemein geredet" (d10).

Darüber hinaus wurde die Skala um *drei neue Items ergänzt*, welche messen sollen, wie häufig sich die Zielschüler/innen kurz vor oder nach dem Unterricht gegenüber anderen Kindern dissozial verhielten (Gemeinheit, Streit, Schlägerei) (Eckstein et al., 2018). Die Idee hinter dieser Ergänzung war, dass solches Verhalten nur sehr selten während der eigentlichen Unterrichtszeit gezeigt wird, aber eine nachwirkende Störung im Unterricht verursachen kann, wenn es z.B. auf dem Schulweg oder in der Pause gezeigt wird.

Schliesslich wurde ein neues, *hybrides Antwortformat* entwickelt. Beim Pretest wurde das erfassbare Maximum mit der obersten Antwortkategorie ("3 Mal und häufiger") eingegrenzt. Um die Auftretenshäufigkeit bei der Hauptstudie differenzierter zu erfassen, enthält die Antwortskala neu 6 vorgegebene Kategorien (0 bis 5 Mal) plus eine Option für handschriftliche Einträge im freien Format ergänzt ("Häufiger, nämlich").

Ein zweites wichtiges Massnahmenpaket der inhaltlichen Überarbeitung des Instrumentariums betraf das Konstrukt des *subjektiven Störungsempfindens*. Beim Pretest bestand ein konzeptuelles Problem in der Operationalisierung der eingesetzten Skalen: Einige Items waren so formuliert, dass sie nicht das Störungsempfinden der Rater gemessen haben, sondern Eigenschaften der Zielschüler/innen (aus Sicht der Rater). Beispielsweise Item (s2) in der Lehrpersonen-Version: "...ist bei den anderen Kindern der Klasse beliebt". Dieses Item erfasste den Beliebtheitsgrad der Zielschüler/innen, welcher aber nur entfernt mit dem Störungsempfinden der befragten Lehrpersonen zu tun hat. Deshalb wurden beide Skalen (affektives, kognitives Störungsempfinden) grundlegend überarbeitet, sodass die Lehrpersonen sowie die Mitschüler/innen aus ihrer *eigenen, subjektiven Perspektive* angeben konnten, inwieweit sie die Zielschüler/innen als störend empfinden (Eckstein et al., 2018).

Zudem wurden die Skalen neu auch für *Selbsteinschätzungen* ausgelegt (beim Pretest wurden diesbezüglich keine Self-Ratings eingeholt). Im Gegensatz zu den beiden anderen Rating Typen sollte damit aber nicht das Störungsempfinden der (Self-)Rater gemessen werden – die Zielschüler/innen sollten einschätzen, inwieweit sie ihre Lehrperson und/oder Mitschüler/innen gestört hatten. Damit wurde die vom Umfeld empfundene Störungsintensität pro Zielschüler/in aus drei Perspektiven erfasst. Als Beispiel mag Item per05 herhalten:

- Self-Rating: Ich habe sie geärgert.
- Peer-Rating: [Zielschüler/in X] hat mich geärgert.
- Teacher-Rating: [Zielschüler/in X] hat mich geärgert.

Eine vollständige Übersicht über alle in der Hauptstudie eingesetzten Items der Konstrukte "Devianz" und "Störungsempfinden" findet sich in Kap. 2.4.1 (Seite 99 f.).

2.3.4. Datencleaning und vorbereitende deskriptive Analysen

Das Datencleaning erfolgte nach einem standardisierten Vorgehen. Beispielsweise wurden Kreuze zwischen zwei Antwortkategorien mittels Münzwurf einer der beiden Kategorien zugewiesen. Mutmassliche Fehler, z.B. identisches Ankreuzen inverser Items, wurden mittels Bildung von Dummyvariablen identifiziert und im Einzelfall unter Bezugnahme weiterer Informationen begutachtet. In einigen wenigen Fällen erschien widersprüchliches Antwortverhalten nicht völlig unplausibel, sodass wir die Angaben unverändert liessen. Dagegen wurden offensichtliche Fehler manuell korrigiert, im Zweifelsfall wurden die Angaben gelöscht.

Bei den Devianz-Items bestand ein Problem darin, dass einige Befragte im freien Format nicht-numerische Angaben notierten (z.B. "immer"). Diese Einträge wurden nachträglich in numerische Werte transformiert (z.B. "immer" = Maximalwert unter den numerischen Angaben) (Eckstein et al., 2018). Darüber hinaus wurde festgestellt, dass einige wenige Angaben von Schülerinnen und Schülern unsinnig hoch waren (z.B. Wert "1000"). Deshalb wurde nach Tabachnick und Fidell (2013) nachträglich eine Obergrenze festgelegt. Hierfür wurde der höchste Wert aller Angaben der Lehrpersonen (60 Mal) herangezogen – höhere Angaben von Kindern wurden auf diesen Maximalwert gekappt. Dies betraf 0.44% aller Selbsteinschätzungen, sowie 0.15% aller Peer-Ratings (Eckstein et al., 2018). Dadurch entstand ein für alle Ratingtypen einheitlicher Wertebereich von 0-60 pro Devianz-Item.

Der bereinigte Datensatz wurde *vorbereitenden, deskriptiven Analysen* unterzogen, um das durchschnittliche Antwortverhalten hinsichtlich der wichtigsten Konstrukte erstmalig zu sichten. Für diese vorläufige Auslegeordnung wurden alle 18 Items des Konstrukts Devianz zu einem korrigierten Summenscore aufaddiert; beim Konstrukt Störungsempfinden wurden alle neun Items zu einem Mean Score gemittelt.

Der Devianz-Summenscore bringt zum Ausdruck, wie oft ein Zielschüler oder eine Zielschülerin alle aufgelisteten Verhaltensweisen insgesamt gezeigt hat. Der mögliche Wertebereich beträgt 0 bis 1'080 (18 Items mal Maximalwert 60). Abbildung 6 zeigt die Summenscores gemäss den Ratings der Lehrpersonen in einem Histogramm.

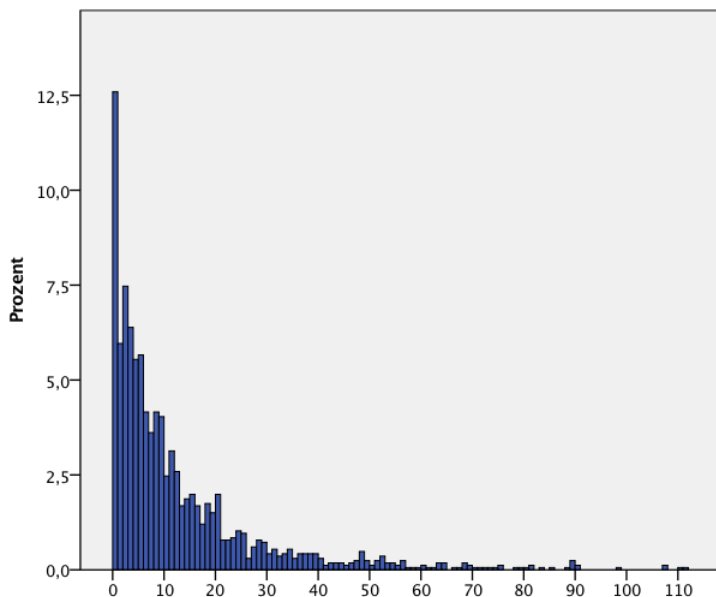


Abbildung 6. Zielkindspezifische Auftretenshäufigkeit unterrichtlicher Devianz gemäss den Lehrpersonen.

Anmerkungen: Die Darstellung ist gekappt oberhalb des 99. Perzentils (113 deviante Verhaltensweisen).

Die Darstellung veranschaulicht, dass die Daten *deutlich rechtsschief* verteilt sind. Der erste Balken zeigt, dass 12.60% der 1'660 Zielschüler/innen nach Einschätzung der Lehrpersonen keine einzige der 18 aufgelisteten Verhaltensweisen während der vorangegangenen zwei Wochen zeigten. Dem Median nach zeigte die Hälfte aller Schüler/innen weniger als 7 deviante Verhaltensweisen; das 95. Perzentil entspricht einem Summenscore von 51.95. Die Peer-Ratings (n = 5'500; Median = 6.00; 95. Perzentil = 38.00) sowie die Selbsteinschätzungen (n = 1'390; Median = 6.00; 95. Perzentil = 30.17) sind ebenfalls deutlich rechtsschief verteilt (Eckstein, Grob & Reusser, 2017).

Ähnliche Bilder ergaben sich auch mit Blick auf das Konstrukt Störungsempfinden. In Abbildung 7 sind die Mean Scores der Teacher-Ratings abgebildet – auch hier zeigt sich eine deutlich rechtsschiefe Verteilung.

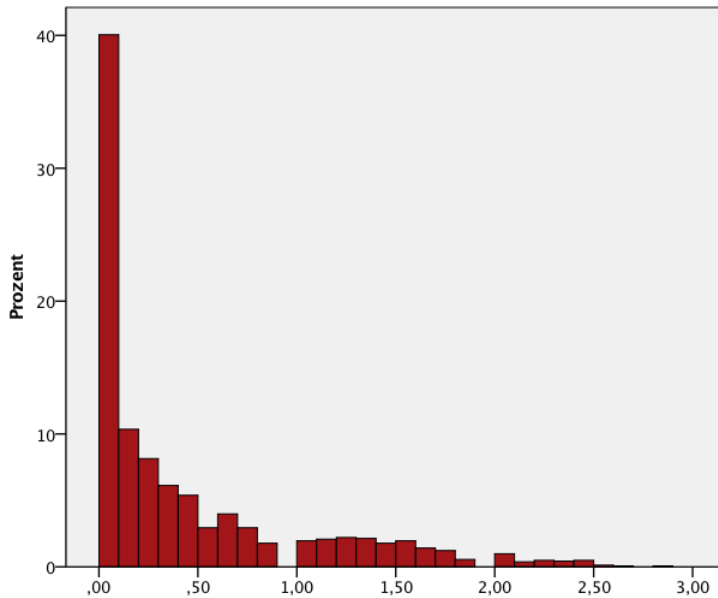


Abbildung 7. Mittleres zielschülerspezifisches Störungsempfinden der Lehrpersonen.

Gemessen am theoretischen Mittelwert der Skala (1.5) werden 91.9 % aller Schüler/innen ($n = 1'633$) von ihren Lehrpersonen als kaum störend beschrieben. In 40.00 % aller Fälle resultiert sogar ein Mean Score von 0.00, welcher besagt, dass die Lehrpersonen diese Kinder als gar nicht störend sondern im Gegenteil als nett und freundlich erleben. Der Median beläuft sich auf 0.11. Wiederum sind die Daten der beiden anderen Rating Typen ähnlich rechtsschief verteilt (Eckstein et al., 2017). Die Peer-Ratings ($n = 5'484$; Median = 0.33) sind zu 88.0% unterhalb des theoretischen Mittels; bei den Selbsteinschätzungen ($n = 1'370$; Median = 0.33) sind sogar 97.2% niedriger als 1.5.

Zwischenfazit: Die Resultate dieser vorbereitenden Analysen deuten darauf hin, dass Unterrichtsstörungen kein flächendeckendes Problem darstellen. Diejenigen Schüler/innen, welche sich häufig deviant verhalten und/oder als störend wahrgenommen werden, sind in der klaren Minderheit. Wie bereits bei der Vorstudie passen auch diese vorläufigen Befunde der Hauptstudie nur bedingt zu einer tendenziell besorgniserregenden, medial verbreiteten Aussensicht auf die (integrative) Schule. Allerdings ist einschränkend festzuhalten, dass die bisherigen Analysen noch sehr allgemein gehalten sind. Es stellt sich daher die Frage, wie sich die Situation auf der Ebene einzelner Schulklassen darstellt.

2.3.5. Perspektivenvergleich innerhalb der an der Studie teilnehmenden Schulklassen

Für jede an der SUGUS-Hauptstudie teilnehmenden Lehrperson wurde ein personalisierter Ergebnisbericht mit Befunden zu ihrer Klasse erstellt (Eckstein, Luger, et al., 2016b). Diese Form der Ergebniskommunikation wurde in Beitrag III der Dissertation (Eckstein, 2018) einem wissenschaftlichen Publikum vorgestellt. In dieser Originalarbeit wurden zwei *Fragestellungen* bearbeitet:

- *Gibt es Perspektivenunterschiede in der Wahrnehmung von Unterrichtsstörungen durch Lehrpersonen gegenüber Schülerinnen und Schülern?*
- *Wie können die Befunde den an der Studie teilnehmenden Lehrpersonen kommuniziert werden?*

Die in der Originalarbeit gegebenen Antworten auf diese Fragen werden im vorliegenden Kapitel 2.3.5 anhand von Beispielen dargelegt und mit zusätzlichen Informationen ergänzt.

Vor dem klassenspezifischen Teil der Ergebnisberichte wurden deskriptive Befunde zur gesamten Stichprobe präsentiert. Dazu wurden die zuvor berichteten absoluten Häufigkeiten der Ratings zu Devianz und Störungsempfinden (Kap. 2.3.4) am Beispiel einer „Durchschnittsklasse“ anschaulich beschrieben. Diese Durchschnittsklasse wurde aus allen 85 an der Studie teilnehmenden Schulklassen errechnet; sie besteht aus 19.85 ($SD = 3.00$) Schülerinnen und Schülern. Aus Sicht der „Durchschnittslehrperson“ zeigten diese knapp 20 Schüler/innen pro Klasse gemeinsam rund 275.56 ($SD = 196.97$) Normabweichungen innert zwei Wochen. Bei 60 Lektionen in zwei Wochen entspricht dies 4.59 Normabweichungen pro Lektion, welche in der Durchschnittsklasse gesamthaft gezeigt wurden. Dabei handelte es sich meistens um eher bagatellhafte Formen wie Schwatzen (Eckstein, Grob, et al., 2015). Darüber hinaus haben die Lehrpersonen ihre Schüler/innen im Mittel auch als wenig störend eingeschätzt: Die von der Durchschnittslehrperson empfundene Störungsintensität pro Zielschüler/in belief sich im Mittel auf 0.43 ($SD = 0.58$).

Im *personalisierten Teil* der Ergebnisberichte wurde die Situation in jeder an der Studie teilnehmenden Schulklassse anhand von vier Kategorien beleuchtet. Dazu wurden vier Variablen gebildet und für die Lehrpersonen (Lp) sowie für die Schülerinnen und Schüler (SuS) ausgewertet.

- 1) Devianz (zielschülerspezifische Teacher-Ratings und Peer-Ratings)
- 2) Störungsempfinden (zielschülerspezifische Teacher-Ratings und Peer-Ratings)
- 3) Allgemeine Sensitivität (raterspezifische Selbsteinschätzung)
- 4) Allgemeine Belastung (raterspezifische Selbsteinschätzung)

Bei den Variablen "Devianz" (1) und "Störungsempfinden" (2) wurden die zielschülerspezifischen Summenscores pro Ratingtyp zu *Klassensummen* aggregiert (die multiplen Peer-Ratings pro Zielschüler/in wurden vorgängig gemittelt). Auf eine Relativierung an der Klassenstärke wurde verzichtet, weil das absolute Problemausmass interessierte, welches durch die kumulierten Werte repräsentiert wird. Aufgrund dieses Verfahrens sind in grossen Klassen theoretisch höhere Werte möglich als in kleinen, was sich in den Resultaten jedoch nicht 1:1 niederschlägt – es gibt z.B. grössere Klassen, in denen alle Schüler/innen zusammen weniger Devianz zeigten als in kleineren Klassen.

Die Variable "Allgemeine Sensitivität" (3) misst, wie empfindlich die Lehrpersonen und die Schüler/innen generell auf Störungen reagieren. Bei der Variable "Allgemeine Belastung" (4) handelt es sich um einen Mischindex, welcher misst, inwieweit die Lehrpersonen und die Schüler/innen den schulischen Alltag als belastend erleben. Die Reliabilität der berechneten Werte²⁰ ist als genügend bis sehr gut zu beurteilen – detaillierte Angaben hierzu finden sich in der Originalarbeit (Eckstein, 2018).

Um den Lehrpersonen die Situation in ihrer Klasse perspektivenvergleichend zu illustrieren, wurden erneut *Netzdiagramme* eingesetzt. Die in den Diagrammen abgebildeten Flächen entsprechen dem wahrgenommenen Problemausmass in der jeweiligen Perspektive (je grösser die Fläche, desto problematischer die Situation). Dazu wurde die unterschiedliche Skalierung der vier Variablen vorgängig mittels *linearer Transformation* vereinheitlicht. Der

²⁰ Wie einleitend beschrieben (Kap. 2.3.1) handelt es sich bei diesen Variablen noch nicht um faktorenanalytisch bestimmte Skalen, dennoch wurde die Qualität der verrechneten Scores empirisch untersucht, etwa mittels Skalenanalysen.

resultierende Wertebereich reicht für alle Variablen von 0 (keine Normabweichung etc.) bis 100, wobei das Maximum dem höchsten in der Stichprobe vorkommenden Wert entspricht. Damit liegt den Endpunkten der Achsen eine soziale Bezugsnorm²¹ zugrunde.

In einem ersten *Netzdiagramm-Typ Lp/Lp* wurden die Angaben der adressierten Lehrperson und die mittleren Angaben aller Lehrpersonen verglichen. In einem zweiten *Diagramm-Typ SuS/SuS* wurde die mittlere Perspektive der Schüler/innen dieser Klasse mit den durchschnittlichen Angaben aller Schüler/innen der Stichprobe verglichen. Schliesslich wurde in einem dritten *Diagramm-Typ Lp/SuS* die Perspektive der adressierten Lehrperson mit der mittleren Perspektive ihrer Schüler/innen verglichen.

Als vorgängige Interpretationshilfe für die klassenspezifischen Grafiken diene ein Diagramm der *Durchschnittsklasse*, welches in Abbildung 8 dargestellt ist. Darin wird die mittlere Perspektive aller Lehrpersonen (blaue Fläche) im Vergleich mit der mittleren Perspektive aller Schüler/innen (gelbe Fläche) dargestellt²². Es handelt sich also um den Perspektivenvergleich Lp/SuS innerhalb der Durchschnittsklasse.

²¹ Wünschbar wäre zwar eine kriteriale Bezugsnorm, doch diese verursacht Darstellungsprobleme: Definiert man beispielsweise die Maxima der theoretisch möglichen Wertebereiche als Endpunkte, fallen die meisten Flächen derart klein aus, dass die Unterschiede zwischen den dargestellten Perspektiven nicht erkennbar sind. Definiert man hingegen einen beliebigen Wert darunter als Endpunkt (z.B. aus theoretischen Überlegungen), überragen die Flächen einiger Klassen die Endpunkte des Diagramms.

²² Damit die Linien zwischen den Achsenabschnitten sichtbar bleiben, wurden die Flächen semitransparent eingefärbt. Dies hatte zur Folge, dass die sich überlagernden Teilflächen zwischen den Angaben der Lehrpersonen (blau) und Schüler/innen (gelb) einen grünlichen Farbton aufweisen.

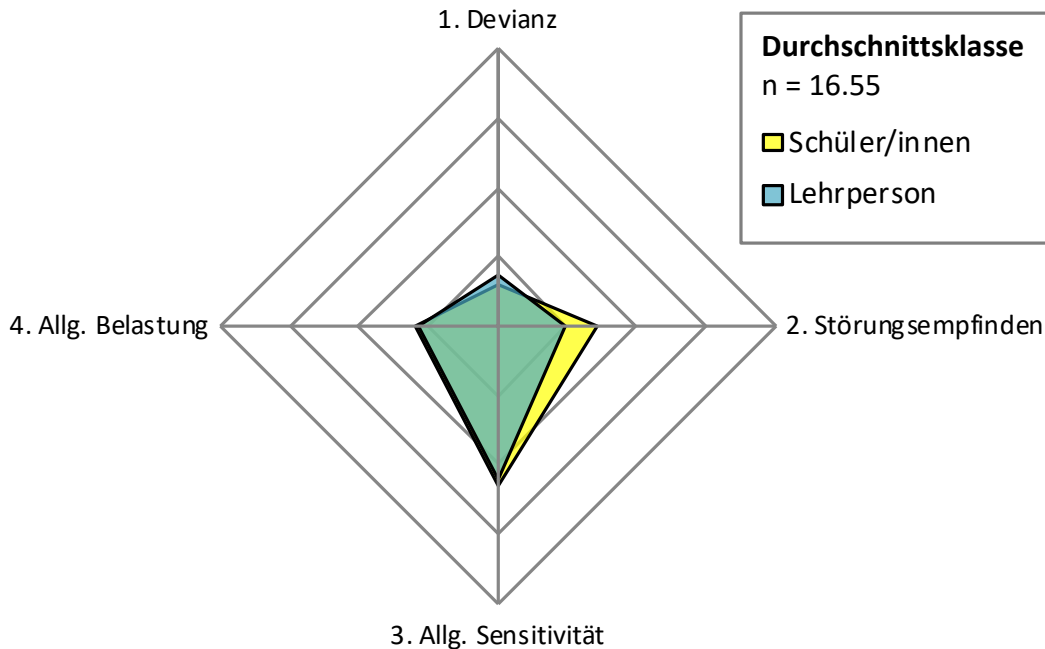


Abbildung 8. Perspektivenvergleich Lp/SuS innerhalb der Durchschnittsklasse.

Die Darstellung illustriert, dass sich die gemittelten Perspektiven der Lehrpersonen und Schüler/innen nur *geringfügig unterscheiden*. Dieser Befund deckt sich weitgehend mit den Resultaten von Varianzanalysen (ANOVAs) und t-Tests, welche auf Stichprobenebene berechnet wurden (Eckstein, 2018). Der eindeutigste Unterschied findet sich bei der Variable Störungsempfinden (2): Eine ANOVA²³ mit Messwiederholung für alle drei Ratertypen fällt signifikant aus [$F(1.9, 2554.7) = 96.3$; $p < .001$; $\eta^2_{\text{part}} = 0,07$]. Bonferroni-korrigierte paarweise Vergleiche zeigen, dass die Zielschüler/innen im Durchschnitt von ihren Peers ($M = .60$; $SD = .47$) als störender beschrieben wurden als von ihrer Lehrperson ($M = .41$; $SD = .58$).

²³ Aufgrund der rechtsschiefen Verteilung der Daten ist die statistische Absicherung der ANOVA möglicherweise nicht zuverlässig. Um das Risiko eines Alpha-Fehlers zu verringern, wurde das Signifikanzniveau bei sämtlichen ANOVAs auf $p < .001$ gesetzt.

Bei manchen *klassenspezifischen Resultaten* zeigten sich jedoch markantere Unterschiede zwischen den raterspezifischen Perspektiven und/oder ein vom Durchschnitt stark abweichendes Problemausmass. Abbildung 9 veranschaulicht diese Befunde anhand klassenspezifischer Netzdiagramme (Lp/SuS) von sechs nicht repräsentativ ausgewählten Klassen, welche sich an der SUGUS Studie beteiligten. Es handelt es sich um interessante Extremfälle, welche sich in mindestens einer Hinsicht deutlich vom Durchschnitt unterscheiden. Die Diagramme werden hier als Ergänzung zu den in der Originalarbeit (Eckstein, 2018) vorgestellten Resultaten präsentiert²⁴.

Bei *Klasse 015* beschrieb die Lehrperson in allen Kategorien ein minimales Problemausmass. Dagegen haben ihre Schüler/innen bei drei von vier Kategorien höhere Werte angegeben als die Schüler/innen der Durchschnittsklasse (diese Differenz lässt sich anhand des abgebildeten Diagramms nicht ablesen – die Schlussfolgerung wurde dem zuvor erwähnten Netzdiagramm des Typs SuS/SuS entnommen). Hat also die Lehrperson untertrieben oder haben ihre Schüler/innen übertrieben? Solchen Fragen wird derzeit in der qualitativen Teilstudie des SUGUS Projekts nachgegangen (Marusic-Würscher & Reusser, 2018).

²⁴ In der Originalarbeit werden zwei fiktive Beispielklassen vorgestellt, welche aus jeweils fünf Klassen gebildet wurden: Bei "Low 5" ist das Problemausmass aus Sicht der Lehrpersonen minimal; bei "Top 5" ist es maximal. Diese zwei Beispielklassen wurden auch in einem anonymisierten Ergebnisbericht (Eckstein, Luger, et al., 2016a) vorgestellt, welcher den an der Studie teilnehmenden Lehrpersonen in Ergänzung zu den für sie personalisierten Berichten (Eckstein, Luger, et al., 2016b) zugestellt sowie anderen interessierten Personen zugänglich gemacht wurde.

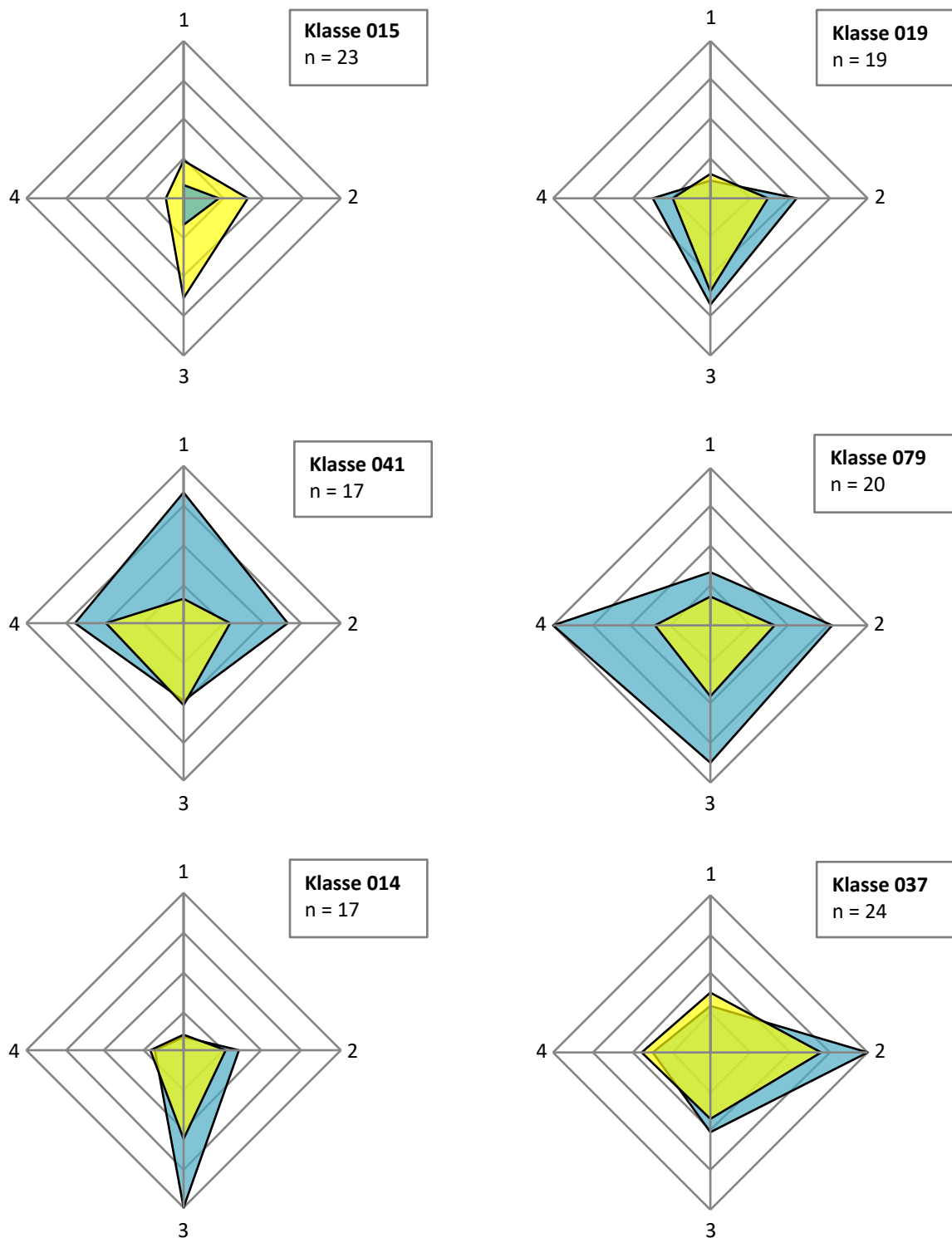


Abbildung 9. Klasseninterne Perspektivenvergleiche Lp/SuS in sechs Klassen.

Anmerkungen: (1) = Devianz; (2) Störungsempfinden; (3) Allgemeine Sensitivität; (4) Allgemeine Belastung; blaue Flächen = Lp; gelbe Fläche = SuS;

Die Lehrperson von *Klasse 019* gab bei der Auftretenshäufigkeit unterrichtlicher Devianz (1) niedrigere Werte an als ihre Schüler/innen (und auch als die Durchschnittslehrperson). Dieses Ergebnis deckt sich mit dem Kommentar der Lehrperson, den diese am Ende des Fragebogens notierte: "Diese Klasse ist untereinander im Umgang sehr liebenswürdig und das Unterrichten mit ihnen sehr angenehm" (Kap. 1). Trotzdem beschreibt die Lehrperson das Problemausmass in den übrigen Kategorien als erhöht verglichen mit ihren Schülerinnen und Schülern (und auch im Vergleich mit der Durchschnittslehrperson).

Nach Auskunft der Lehrperson von *Klasse 041* zeigten deren Schüler/innen gemeinsam 1'174 Normabweichungen in den zwei Wochen vor der Befragung (1). In keiner anderen Klasse wurden höhere Werte in Bezug auf diese Variable berichtet²⁵. Entsprechend berichtete die Lehrperson auch hohe Werte beim Störungsempfinden (2) sowie bei der Allgemeinen Belastung (4). Dagegen schildern die Schüler/innen von Klasse 041 das Problemausmass im Vergleich zu ihrer Lehrperson als weniger gravierend. Doch verglichen mit den Schüler/innen der Durchschnittsklasse beschreiben auch sie sich im Mittel als stärker belastet (4).

²⁵ Gemäss den vorherigen Anmerkungen zur Vereinheitlichung der Wertebereiche der vier Variablen wurden auf der Grundlage dieses Maximalwerts ($1'174 = 100$) alle Werte dieser Variable mittels Dreisatz linear transformiert. Dass die blaue Fläche von Lp 041 im Diagramm trotzdem nicht bis zum Endpunkt der Achse reicht, liegt daran, dass für diesen Perspektivenvergleich nur Angaben zu SuS beigezogen werden konnten, welche sich an der Studie beteiligten (nicht teilnehmende SuS wurden von den Peers nicht eingeschätzt). Da sich zwei SuS der Klasse 041 von der Studie abgemeldet hatten, wurden ihre "Beiträge" zur Klassensumme gemäss Lp vor dem Perspektivenvergleich wieder abgezogen.

Auch in *Klasse 079* scheint das Problemausmass aus Sicht der Lehrperson sehr hoch: Sie hat bei allen vier Variablen höhere Werte angegeben als die Durchschnittslehrperson, bei der Allgemeinen Belastung (4) hat sie gar den höchsten Wert in der Stichprobe (2.29 bei einem theoretischen Maximum von 3.0). Diese Lehrperson ist nach eigener Auskunft ziemlich nahe an einem Burnout. Demgegenüber sind die Angaben ihrer Schüler/innen deutlich niedriger und auch im Vergleich mit der Durchschnittsklasse nur geringfügig erhöht.

Bei *Klasse 014* fällt auf, dass sich die Lehrperson trotz hoher Sensitivität (3) als kaum belastet beschreibt (4), vermutlich weil sich die Schüler/innen in dieser Klasse vergleichsweise selten deviant verhalten (1). Interessanterweise beschreibt diese Lehrperson ihre Schüler/innen trotzdem als etwas störender (2) – sowohl im Vergleich mit den Peer-Ratings als auch verglichen mit der Durchschnittslehrperson. Bei der Nachbereitung der personalisierten Ergebnisberichte fiel auf, dass dieses Muster in ähnlicher Form in mehreren Klassen auffindbar ist. Womöglich ist dies darauf zurückzuführen, dass diese Lehrpersonen aufgrund ihrer vergleichsweise erhöhten Sensitivität bereits erste Anflüge devianten Schülerverhaltens registrieren und darauf reagieren, sodass sie viele der sich anbahnenden Probleme frühzeitig eindämmen können. Das sind aber nur Vermutungen und müssten noch geprüft werden.

In der *Klasse 037* beschreiben sich weder die Schüler/innen noch die Lehrperson als besonders empfindlich (3), beurteilen das Problemausmass in den anderen drei Kategorien im Vergleich zur Durchschnittsklasse aber als erhöht. Insbesondere werden die Schüler/innen in dieser Klasse als störender (2) beschrieben als im Durchschnitt, wobei das Störungsempfinden der Lehrperson noch etwas höher ausfällt als dasjenige der Schüler/innen.

2.3.6. Grenzen des deskriptiven Perspektivenvergleichs

Die globalen Mittelwertsvergleiche ergaben nur wenig nennenswerte Unterschiede zwischen der durchschnittlichen Perspektive der Lehrpersonen und derjenigen der Schüler/innen an. Diese vorläufige Feststellung erstaunt mit Blick auf die theoretisch angeführten Gründe, warum Einschätzungen zu Unterrichtsstörungen immer einen subjektiven Anteil einschliessen (Kap. 2.2.5). Die *unerwartet geringen Divergenzen* dürften deshalb zustande gekommen sein, weil die multiplen Peer-Ratings pro Zielschüler/in vor dem Vergleich mit den Teacher-Ratings gemittelt wurden, sodass viel Varianz auf der Ebene der unigen Peer-Ratings verloren ging. Deshalb wurde erwartet, dass sich bei mehrbenenanalytischer Berücksichtigung der Individualebene der Peer-Ratings deutlichere Divergenzen zu den Teacher-Ratings zeigen.

Eine weitere Ursache der geringen Mittelwertsdifferenzen könnte darin liegen, dass die Konstrukte in den bisherigen Analysen ohne Berücksichtigung ihrer faktoriellen Struktur eindimensional untersucht wurden. Die nachträglichen Faktorenanalysen haben jedoch ergeben, dass die Konstrukte "Devianz" und "Störungsempfinden" jeweils zwei Subdimensionen enthalten (Eckstein et al., 2018). Aufgrund der zusammenfassenden Berechnung zweier Subdimensionen ging vermutlich raterspezifische Varianz verloren. Daher wurde angenommen, dass der Perspektivenvergleich auf der Grundlage der faktorenanalytisch bestimmten Dimensionalität der Konstrukte präziser und daher deutlicher ausfallen würde.

Die Analysen der absoluten Häufigkeiten ergaben, dass die durchschnittliche Belastung der untersuchten Stichprobe eher gering ausgeprägt war: *Die meisten Schüler/innen zeigten (sehr) selten deviantes Verhalten und wurden entsprechend als wenig oder gar nicht störend beurteilt.* Die klassenspezifischen Analysen zeigen jedoch, dass die Belastung in einzelnen

Klassen sehr wohl stark ausgeprägt war. Als Beispiel für eine eine Klasse mit erhöhter Belastung mag Klasse 041 herhalten, deren Netzdiagramm Lp/SuS auch in Abbildung 9 enthalten ist. Wie schwierig es in dieser Situation für alle Beteiligten gewesen sein musste, belegen einige Kommentare, welche die Lehrperson und einige Schüler/innen dieser Klasse am Ende des Fragebogens notierten:

Ich habe die Stelle am 1.8.15 angetreten und per 31.7.16 gekündigt, da eine Klassenintervention nichts brachte und die Schulleitung nicht hinter mir steht. [...] Diese Klasse wurde aus einer 13er und einer 7ner Klasse zusammengelegt. Die frühere KLP der 13er Klasse legte ihr Veto ein, da die Klasse kaum führbar sei, da alle "Untragbaren" mit einigen "Problemlosen" in den letzten 4 Jahren in die 13er Klasse umgesiedelt wurden. [...] Die Drahtzieher dieser Klasse hatten beschlossen, die "neue" KLP wegzuekeln, egal wer es ist. Sie wollten (und wollen heute noch) einen "Herrn [...]", der mal hier im Praktikum war. Es wurden verschiedene Geldbeträge, Esswaren, Bücher, Schulmaterial... gestohlen. Interventionen nützten kaum. Grundsätzliche Regeln werden trotz versch. Massnahmen nicht eingehalten. Eltern und SL unternehmen nichts.
(Klasse 041, Lehrperson)

Ich bin nur von meiner Lehrerin gestört sonst nie! (Klasse 041, Schüler/in 09)

Ich finde es blöd das Lehrperson manchmal die Kinder zusammenscheist obwohl sie gar nichts gemacht haben. (Klasse 041, Schüler/in 07)

Es nervt mich wenn andere Kinder über die Lehrperson oder über andere Kinder schlecht reden. Wenn die Lehrperson etwas sagt und die anderen Kindern das nicht cool finden, schimpfen sie in der Pause über die Lehrperson. (Klasse 041, Schüler/in 17)

Angesichts der sonstigen Resultate ist die Situation in Klasse 041 aber als Extremform zu beurteilen. Darüber hinaus deutet der Einzelbefund aber an, wie stark die jeweiligen Perspektiven sich doch unterscheiden können. Ausgehend von solchen einzelnen, klassenspezifischen Abweichungen von den mittleren Befunden der gering belasteten "Durchschnittsklasse" wurde der Perspektivenvergleich in Beitrag IV der Dissertation (Eckstein, under review) mehrebenenanalytisch untersucht.

2.4. Beitrag IV: Konsistenz und Spezifität der verschiedenen Perspektiven

Abstract aus der Originalarbeit

Classroom disturbances impair the quality of teaching and learning, and they can be a source of strain for both teachers and students. Some studies indicate, however, that not everyone involved gets equally disturbed by the same occurrences. Altogether, there is still little solid knowledge about the teachers' and the students' subjective perception of disturbance. Moreover, rater effects may have confounded the findings available. Addressing these desiderata, the SUGUS study investigates two elements of classroom disturbances within an interactionist framework: the incidence of deviant behaviour shown by particular target students, and the intensity of disturbance as subjectively perceived by teachers, by classmates, and by the targets themselves. For this purpose, we conducted a questionnaire survey among 85 primary-school class teachers and 1412 students. The data were analysed by means of a two-level correlated trait – correlated method minus one [CT-C(M-1)] model. This relatively novel statistical procedure has only rarely been applied in educational research so far. It made it possible to determine the respondents' common view on classroom disturbances as well as the rater-specific perspectives. The results indicate that increasing deviance coincides with increasing distraction and annoyance – but mainly in a relatively small intersection of the different perspectives. Beyond that, the analysis revealed substantial rater effects which explain 30 to 61% of variance in teacher ratings, for instance. The author discusses likely reasons why disturbances are perceived so divergently. (Eckstein, under review)²⁶

Die in Beitrag IV unternommenen Analysen beruhen auf einem vorgängigen Skalierungsverfahren, welches in Kapitel 2.4.1 skizziert wird. Im Anschluss daran werden varianzanalytisch bestimmte Perspektivendifferenzen vorgestellt (Kap. 2.4.2), um die im Beitrag ermittelten Resultate des CT-C(M-1) Modells unter deskriptiven Gesichtspunkten verständlich einordnen zu können. Zentrale Ergebnisse dieses Modells werden in Kapitel 2.4.3 präsentiert.

²⁶ Die vollständigen bibliographischen Angaben der Originalarbeit lauten: Eckstein, B. (under review). Production and Perception of Classroom Disturbances – A new approach to investigating the perspectives of teachers and students.

2.4.1. Skalierung der Konstrukte Devianz und Störungsempfinden in je zwei Dimensionen

Wie in Kap. 2.3.1 erwähnt, wurden die zentralen Konstrukte zuerst deskriptiv ausgewertet, um den an der Hauptstudie teilnehmenden Lehrpersonen den versprochenen Ergebnisbericht zeitnah zu erstellen (Eckstein, Luger, et al., 2016a, 2016b). Die *faktorenanalytisch begründete Skalierung* der Konstrukte erfolgte erst danach. Aus diesen Gründen unterscheiden sich die in Beitrag III der Dissertation (Eckstein, 2018) vorgestellten Konstrukte von denjenigen, welche in Beitrag IV (Eckstein, under review) analysiert wurden.

Der wesentliche Unterschied besteht darin, dass die Konstrukte "Devianz" und "Störungsempfinden" *neu jeweils zweidimensional* gebildet wurden, während sie in Beitrag III ohne Berücksichtigung der faktoriellen Struktur eindimensional gebildet wurden. Bei Devianz ergaben die Faktorenanalysen folgende zwei Dimensionen: undiszipliniertes Verhalten; dissoziales Verhalten. Eine vollständige Übersicht über die 18 eingesetzten Items findet sich in Tabelle 1. Beim Störungsempfinden wurden wie beim Pretest (Kap. 2.2.9) zwei Dimensionen ermittelt: affektives Störungsempfinden; kognitives Störungsempfinden. Eine Übersicht über die Items findet sich in Tabelle 2.

Die Skalierung erfolgte mit Blick auf mehrere Kriterien, u.a. sollten *über alle Ratingtypen hinweg identische, theoretisch plausible, eindimensionale Konstrukte* gebildet werden – unter Berücksichtigung statistischer Konventionen (z.B. interne Konsistenz). Eine vollständige Dokumentation sämtlicher Kriterien sowie eine detaillierte Darstellung der Analyseschritte findet sich im technischen Bericht (Eckstein et al., 2018).

Tabelle 1. Unterrichtliche Devianz – Übersicht über alle Items.

Undiszipliniertes Verhalten			
Item	Wording Self-Ratings	Wording Peer-Ratings	Wording Teacher-Ratings
dev01	Ich habe im Unterricht nicht richtig mitgearbeitet sondern etwas anderes gemacht.	Hat im Unterricht nicht richtig mitgearbeitet sondern etwas anderes gemacht.	Hat im Unterricht nicht richtig mitgearbeitet sondern etwas anderes gemacht.
dev02	Ich habe im Unterricht ohne Aufstrecken etwas gesagt, obwohl man aufstrecken musste.	Hat im Unterricht ohne Aufstrecken etwas gesagt, obwohl man aufstrecken musste.	Hat im Unterricht ohne Aufstrecken etwas gesagt, obwohl das verlangt war.
dev03	Ich habe im Unterricht mit einem anderen Kind geredet, obwohl wir ruhig sein sollten.	Hat im Unterricht mit einem anderen Kind geredet, obwohl es ruhig sein sollte.	Hat im Unterricht mit einem anderen Kind geredet, obwohl es ruhig sein sollte.
dev04	Ich habe im Unterricht Lärm gemacht.	Hat im Unterricht Lärm gemacht.	Hat im Unterricht Lärm gemacht.
dev05	Ich habe der Lehrperson dazwischen geredet.	Hat der Lehrperson dazwischen geredet.	Hat mir dazwischen geredet.
dev06	Ich habe nicht sofort gehorcht, als die Lehrperson etwas sagte.	Hat nicht sofort gehorcht, als die Lehrperson etwas sagte.	Hat eine Anweisung von mir nicht sofort befolgt.
dev07	Ich habe wütend mit der Lehrperson geredet.	Hat wütend mit der Lehrperson geredet.	Hat wütend mit mir geredet.
dev08	Ich habe der Lehrperson freche Antworten gegeben.	Hat der Lehrperson freche Antworten gegeben.	Hat mir freche Antworten gegeben.
Dissoziales Verhalten			
Item	Wording Self-Ratings	Wording Peer-Ratings	Wording Teacher-Ratings
dev09	Ich habe im Unterricht ein anderes Kind ausgelacht.	Hat im Unterricht ein anderes Kind ausgelacht.	Hat im Unterricht ein anderes Kind ausgelacht.
dev10	Ich habe im Unterricht über ein anderes Kind gemein geredet.	Hat im Unterricht über ein anderes Kind gemein geredet.	Hat im Unterricht über ein anderes Kind gemein geredet.
dev11	Ich habe im Unterricht ein anderes Kind beleidigt.	Hat im Unterricht ein anderes Kind beleidigt.	Hat im Unterricht ein anderes Kind beleidigt.
dev12	Ich habe im Unterricht Sachen von einem anderen Kind auf den Boden geschmissen.	Hat im Unterricht Sachen von einem anderen Kind auf den Boden geschmissen.	Hat im Unterricht Sachen von einem anderen Kind auf den Boden geschmissen.
dev13	Ich habe im Unterricht ein anderes Kind angeschrien.	Hat im Unterricht ein anderes Kind angeschrien.	Hat im Unterricht ein anderes Kind angeschrien.
dev14	Ich habe im Unterricht ein anderes Kind geschubst.	Hat im Unterricht ein anderes Kind geschubst.	Hat im Unterricht ein anderes Kind geschubst.
dev15	Ich habe im Unterricht ein anderes Kind geschlagen oder getreten.	Hat im Unterricht ein anderes Kind geschlagen oder getreten.	Hat im Unterricht ein anderes Kind geschlagen oder getreten.
dev16	Ich war gemein zu einem anderen Kind (z.B. Gerüchte verbreiten, Mobbing).	War gemein zu einem anderen Kind (z.B. Gerüchte verbreiten, Mobbing).	War gemein zu einem anderen Kind (z.B. Gerüchte verbreiten, Mobbing).
dev17	Ich habe mit einem anderen Kind gestritten (z.B. beleidigen, anschreien).	Hat mit einem anderen Kind gestritten (z.B. beleidigen, anschreien).	Hat mit einem anderen Kind gestritten (z.B. beleidigen, anschreien).
dev18	Ich habe ein anderes Kind geschlagen oder getreten.	Hat ein anderes Kind geschlagen oder getreten.	Hat ein anderes Kind geschlagen oder getreten.

Anmerkungen: Die Items wurden zielschülerspezifisch beantwortet. Es wurde erfasst, wie häufig die Zielschüler/innen die 18 Verhaltensweisen in der 2 Wochen vor der Befragung gezeigt hatten. Dazu nutzten die Rater ein hybrides Antwortformat mit sechs vorgegebenen Antwortkategorien ("Nie" = 0 bis "5 Mal" = 5) plus einer Option für freie Antworten ("Häufiger, nämlich). Weiterführende Angaben (z.B. Mantelfrage, Trennschärfen, Faktorladungen) finden sich im technischen Bericht (Eckstein et al., 2018).

Tabelle 2. Subjektives Störungsempfinden – Übersicht über alle Items.

Affektives Störungsempfinden			
Item	Wording Self-Ratings	Wording Peer-Ratings	Wording Teacher-Ratings
per01r	Ich war immer nett zu ihnen. (r)	...war immer nett zu mir. (r)	...war immer nett zu mir. (r)
per04	Ich ging ihnen auf die Nerven.	...ging mir auf die Nerven.	...ging mir auf die Nerven.
per05	Ich habe sie geärgert.	...hat mich geärgert.	...hat mich geärgert.
per09r	Ich war immer freundlich zu ihnen. (r)	...war immer freundlich zu mir. (r)	...war immer freundlich zu mir. (r)
Kognitives Störungsempfinden			
Item	Wording Self-Ratings	Wording Peer-Ratings	Wording Teacher-Ratings
per02	Ich habe den Unterricht aufgehalten.	...hat mich im Unterricht vom Lernen abgehalten.	...hat mich vom eigentlichen Unterrichten abgehalten.
per03	Ich habe den Unterricht gestört.	...hat mich im Unterricht gestört.	...hat mich im Unterricht gestört.
per06	Ich habe ihre Konzentration gestört.	...hat meine Konzentration gestört.	...hat meine Konzentration gestört.
per07	Ich habe sie im Unterricht abgelenkt.	...hat mich im Unterricht abgelenkt.	...hat mich im Unterricht abgelenkt.

Anmerkungen: Die Items wurden zielschülerspezifisch mittels vierstufigem Antwortformat beantwortet ("Stimmt gar nicht" = 0 bis "Stimmt genau" = 3). Die Items per01r und per09r waren positiv formuliert, damit die Zielschüler/innen nicht durchweg negativ beschrieben werden mussten (die Antworten wurden nachträglich invertiert).

Die Self-Ratings der Zielschüler/innen waren im Fragebogen nach den Peer-Ratings platziert, weil wir davon ausgegangen sind, dass es den Kindern so leichter fallen würde, die von ihnen ausgehende Störungsintensität ohne Verzerrung aufgrund selbstwertdienlicher Antwortstrategien einzuschätzen. Die Mantelfrage dazu lautete: "Und wie warst Du zu deinen Mitschülern und zu deiner Lehrperson?".

2.4.2. Varianzanalytischer Perspektivenvergleich anhand der vier Konstrukte

Im nachfolgenden Kapitel 2.4.3 wird ein komplexes Strukturgleichungsmodell vorgestellt (Abbildung 14, S. 110). Um die Resultate dieses SEM besser einordnen zu können, werden im vorliegenden Abschnitt die Ergebnisse vorbereitender Analysen präsentiert. Es handelt sich um Mittelwertsvergleiche der drei Ratingtypen, welche in Säulendiagrammen zusammenfassend dargestellt werden.

Zur Vermeidung der α -Fehlerinflation (Bühner & Ziegler, 2009) wurden nicht multiple t-Tests, sondern *ANOVAs mit Messwiederholung* berechnet²⁷. Dabei stellen die Self-Ratings der Zielschüler/innen sowie die Teacher-Ratings und die mittleren Peer-Ratings pro Zielschüler/in jeweils eine Messung der interessierenden Eigenschaft dar.

Liegen die mittleren Einschätzungen eines Ratertyps höher als bei den anderen Typen, handelt es sich technisch gesprochen um eine durchschnittliche *Überschätzung* bei diesem einen Typus – oder um *Unterschätzungen* durch die anderen. Aufgrund der multiplen Interpretationsmöglichkeiten werden die *mittleren Peer-Ratings als Referenzmethode* herangezogen.

Die Abweichungen der anderen Ratertypen von dieser Referenzmethode werden in einem rein technischen Sinne als Über- oder Unterschätzung bezeichnet. Damit soll nicht insinuiert werden, dass die Peer-Ratings objektiv richtig liegen. Die Angemessenheit der verschiedenen Perspektiven wird später noch diskutiert. Bei der schriftlichen Ergebnisdarstellung in diesem Abschnitt liegt der Fokus auf dem Vergleich der Teacher-Ratings mit den mittleren Peer-Ratings.

²⁷ Wie bereits erwähnt (Kap. 2.3.4), ist die statistische Absicherung der ANOVAs aufgrund der rechtsschiefen Verteilung der Daten möglicherweise nicht zuverlässig. Ein Problem könnte insbesondere darin liegen, kleine aber signifikante Effekte irrtümlich als reale Unterschiede zu deuten (Fehler 1. Art). Aufgrund der Höhe der ermittelten Effekte kann jedoch davon ausgegangen werden, dass diese tatsächlich in der Population vorkommen und nicht nur zufällig in der vorliegenden Stichprobe gefunden wurden. Dennoch wurde das Signifikanzniveau auf $p < .001$ gesetzt, um das Risiko des α -Fehlers zu verringern.

Eine erste ANOVA wurde für das Konstrukt *undiszipliniertes Verhalten* berechnet. Der Haupteffekt für den Vergleich der drei Perspektiven fiel signifikant aus, mit geringer Effektstärke [$F(1.49, 2046.56) = 76.08; p < .001; \eta^2_{\text{part}} = .05$]. Die Innersubjektkontraste und Bonferroni-korrigierten paarweisen Vergleiche zeigen, dass sich einzig die Einschätzungen der Lehrpersonen im Vergleich zu den zwei anderen Ratingtypen signifikant unterscheiden: Verglichen mit den mittleren Peer-Ratings ($M = .84; SD = .82$) haben die Lehrpersonen ($M = 1.31; SD = 2.05$) die Zielschüler/innen im Durchschnitt als undisziplinierter beschrieben, bei mittlerer Effektstärke [$F(1, 1376) = 106.87; p < .001; \eta^2_{\text{part}} = .07$]. Abbildung 10 illustriert diese Ergebnisse zusammenfassend.

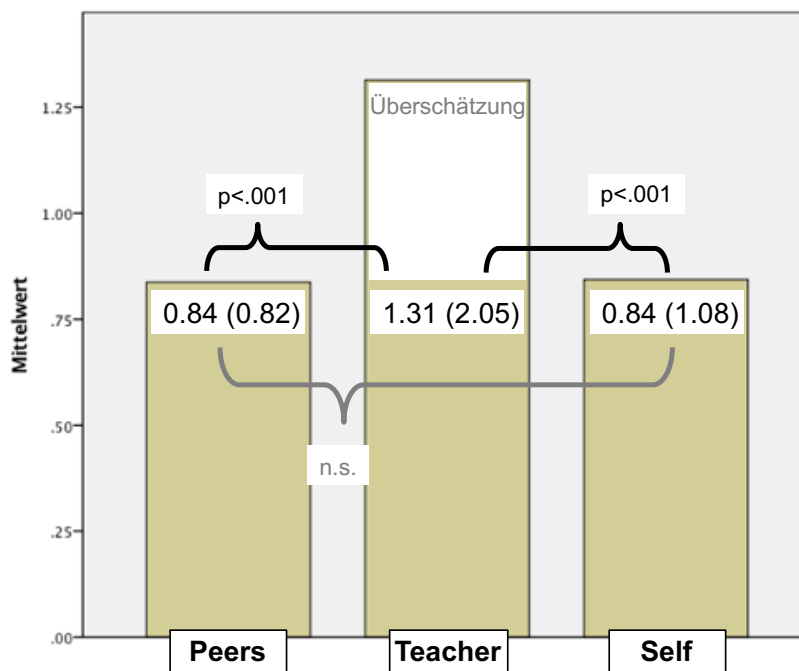


Abbildung 10. Varianzanalytischer Perspektivenvergleich des Konstrukts *undiszipliniertes Verhalten*.

Anmerkungen (gelten für die Abbildungen 10–13):

Die Höhen der Säulen illustrieren die Mean Scores pro Ratingtyp, welche zusätzlich in den weiss hinterlegten Feldern abgedruckt sind (daneben in Klammern: Standardabweichung). Die geschweiften Klammern (Akkoladen) illustrieren die Bonferroni-korrigierten paarweisen Vergleiche. Signifikante Abweichungen von den mittleren Peer-Ratings (Referenzmethode) werden mit weissen Flächen gekennzeichnet und als Über- bzw. Unterschätzung bezeichnet.

Eine zweite ANOVA wurde für das Konstrukt *dissoziales Verhalten* berechnet. Der Haupteffekt für den Vergleich der drei Perspektiven fiel signifikant aus, mit geringer Effektstärke [$F(1.84, 2528.65) = 29.33; p < .001; \eta^2_{\text{part}} = .02$]. Gemäss den Innersubjektkontrasten und Bonferroni-korrigierten paarweisen Vergleichen geht der Haupteffekt im Wesentlichen daraus hervor, dass die Lehrpersonen im Durchschnitt ($M = .23; SD = .59$) niedrigere Werte angegeben haben als die Peers ($M = .40; SD = .54$) – mit einer mittleren Effektstärke [$F(1, 1376) = 106.87; p < .001; \eta^2_{\text{part}} = .07$]. Die Resultate finden sich zusammenfassend in Abbildung 11.

Folglich unterschätzen die Lehrpersonen die Auftretenshäufigkeit dissozialen Verhaltens in der Tendenz, während sie die Undiszipliniiertheit ihrer Schüler/innen im Mittel überschätzen (verglichen mit den mittleren Peer-Ratings). Der Rater-Effekt geht also bei den zwei Konstrukten in die jeweils entgegengesetzte Richtung. Dies lässt sich vermutlich mit rollenspezifischen Voraussetzungen erklären (Wettstein et al., 2018) – z.B. weil Lehrpersonen Undiszipliniiertheit sensibler registrieren als Schüler/innen und/oder weil Dissozialität oft verdeckt gezeigt wird.

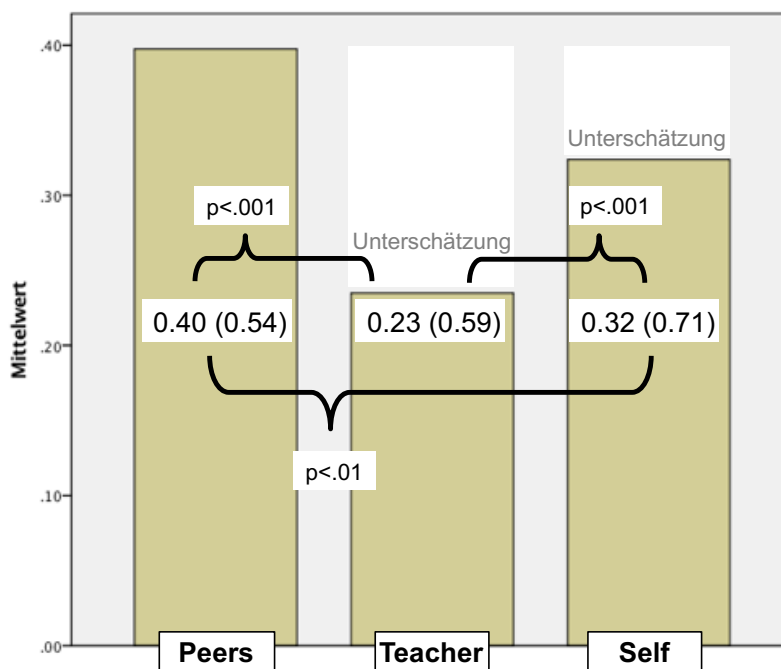


Abbildung 11. Varianzanalytischer Perspektivenvergleich des Konstrukts *dissoziales Verhalten*.

Eine dritte ANOVA wurde für das Konstrukt *affektives Störungsempfinden* berechnet. Der Haupteffekt fiel mit grosser Effektstärke signifikant aus [$F(1.97, 2639.96) = 331.44; p < .001; \eta^2_{\text{part}} = .20$]. Die Innersubjektkontraste zeigen, dass v.a. die Ratings der Peers im Mittel ($M = .69$; $SD = .50$) höher ausfallen als die anderen Ratings. Gemäss paarweisem, Bonferroni-korrigierten Vergleich beurteilten die Peers die Zielschüler/innen im Durchschnitt als störender als die Lehrpersonen ($M = .29$; $SD = .48$) – diese Differenz ist statistisch höchst bedeutsam, bei einer grossen Effektstärke [$F(1, 1339) = 751.94; p < .001; \eta^2_{\text{part}} = .36$].

Die Resultate finden sich zusammenfassend in Abbildung 12. Mit diesen Ergebnissen wird eine Tendenz repliziert, welche bereits beim Pretest festgestellt wurde (Kap. 2.2.10): Während die Schüler/innen durchaus berichten, dass sie sich über einzelne Mitschüler/innen ärgern oder diese als unfreundlich erleben, sind die Lehrpersonen sehr zurückhaltend mit solchen Zuschreibungen.

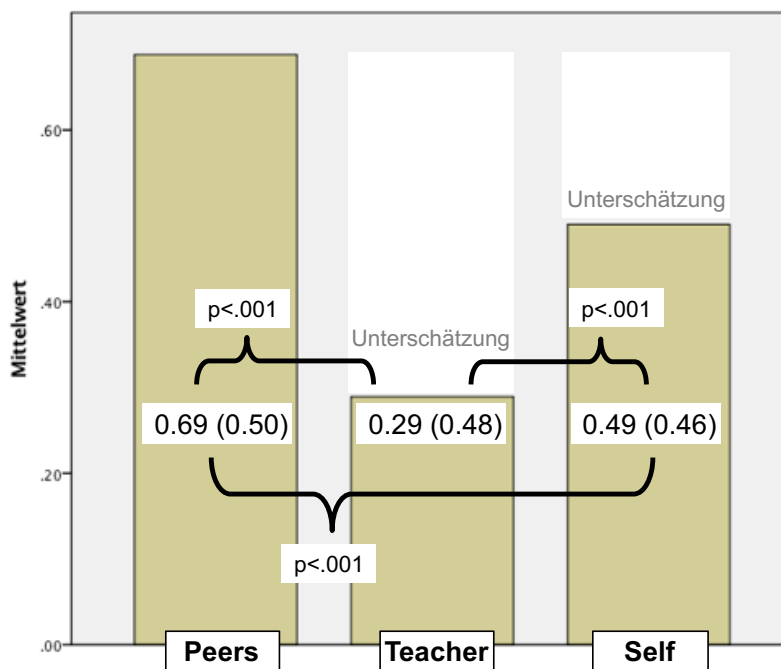


Abbildung 12. Varianzanalytischer Perspektivenvergleich des Konstrukts *affektives Störungsempfinden*.

Schliesslich wurde eine vierte ANOVA für das Konstrukt *kognitives Störungsempfinden* berechnet. Der Haupteffekt fiel signifikant aus, allerdings mit vernachlässigbarer Effektstärke [$F(1.87, 2509.20) = 14.19; p < .001; \eta^2_{\text{part}} = .01$]. Die paarweisen Vergleiche zeigen, dass sich die durchschnittlichen Einschätzungen der Lehrpersonen ($M = .50; SD = .74$) nicht von den mittleren Peer-Ratings ($M = .51; SD = .48$) unterscheiden [$F(1, 1337) = .17; p = .68$]. Einzig die Self-Ratings fallen im Durchschnitt ($M = .42; SD = .53$) etwas niedriger aus als die anderen beiden Ratingtypen [$F(1, 1337) = .26.66; p < .001$], wobei auch diese Differenz mit Blick auf die Effektstärke ($\eta^2_{\text{part}} = .02$) nicht überbewertet werden sollte. Die Resultate finden sich zusammenfassend in Abbildung 13.

Während somit das affektive Störungsempfinden höchst subjektiv ausfällt, schätzen es die Rater im Durchschnitt weitgehend ähnlich ein, inwieweit die Zielschüler/innen ihr Umfeld ablenken oder in der Konzentration beeinträchtigen.

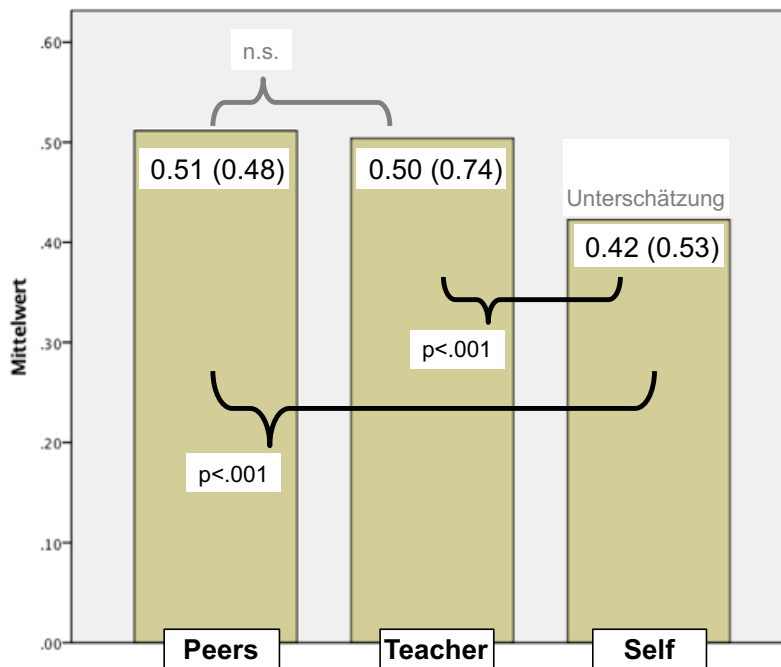


Abbildung 13. Varianzanalytischer Perspektivenvergleich des Konstrukts *kognitives Störungsempfinden*.

Die in diesem Kapitel 2.4.2 präsentierten Mittelwertsdifferenzen bringen zum Ausdruck, inwieweit sich die Rater typen bezogen auf die einzelnen Konstrukte *im Durchschnitt* voneinander unterscheiden. Diese mittleren Differenzen sagen im Prinzip noch nichts darüber aus, wie konsistent bzw. inkonsistent die Rater die einzelnen Zielschüler/innen eingeschätzt haben. Es könnte zum Beispiel sein, dass sich substantielle Divergenzen zwischen den verschiedenen Ratingtypen pro Zielschüler/in nicht in signifikanten Mittelwertsdifferenzen niedergeschlagen haben, weil sich einzelne Über- und Unterschätzungen im Mittelwert über alle Zielschüler/innen gegenseitig "neutralisiert" haben.

Diesen bislang ungeklärten *Grad der Übereinstimmung* (Konsistenz) zwischen den verschiedenen Ratingtypen thematisiere ich im anschliessenden Kapitel 2.4.3. Darin stelle ich ein Strukturgleichungsmodell vor, welches zum einen eine gruppenübergreifend intersubjektiv geteilte Sichtweise auf die Konstrukte ermittelt. Zum anderen schätzt das Modell raterspezifische Abweichungen von dieser gemeinsamen Sichtweise (Rater-Effekte).

2.4.3. Perspektivenvergleich anhand des CT-C(M-1) Ansatzes

Ein erstes Ziel von Beitrag IV der Dissertation (Eckstein, under review) bestand darin, den theoretisch postulierten objektiven Kern von Unterrichtsstörungen – die Auftretenshäufigkeit unterrichtlicher Devianz – approximativ zu bestimmen. Ein zweites Ziel bestand darin, die Relationen zwischen (objektiver) Devianz und (subjektivem) Störungsempfinden zu untersuchen. Die Basis dafür wurde mit der multiperspektivisch angelegten Befragung (Kap. 2.2.7, 2.3.2) anhand des neu entwickelten Instrumentariums (Kap. 2.2.8, 2.3.3) geschaffen, wonach vier skalierte Konstrukte (Kap. 2.4.1) analysiert werden konnten.

In der (englischsprachigen) Originalarbeit bin ich vier *Fragestellungen* nachgegangen, die sich wie folgt sinngemäss auf Deutsch übersetzen lassen:

- (F-1) *Zu welchem Anteil sind die Teacher-, Peer- und Self-Ratings konsistent in Bezug auf die Häufigkeit der von den Zielschülerinnen und Zielschülern gezeigten Devianz und zu welchem Anteil sind die Einschätzungen raterspezifisch?*
- (F-2) *Zu welchem Anteil sind die Teacher-, Peer- und Self-Ratings konsistent in Bezug auf die von den Zielschülerinnen und Zielschülern ausgelöste Störungsintensität und zu welchem Anteil sind die Einschätzungen raterspezifisch?*
- (F-3) *Wie stark korreliert die Häufigkeit devianten Verhaltens der Zielschüler/innen mit der von ihnen ausgelösten Störungsintensität in der von allen Ratern intersubjektiv geteilten Sichtweise?*
- (F-4) *Lassen sich die raterspezifischen Abweichungen von der geteilten Sichtweise über mehrere Konstrukte generalisieren?*

Diesen Fragen bin ich mittels eines Strukturgleichungsmodell des Typs "*Correlated Trait – Correlated Method Minus One*" [CT-C(M-1)] nachgegangen (Eid et al., 2003). Dabei handelt es sich um ein spezielles Multitrait-Multimethod-(MTMM)-Verfahren (Geiser, Eid, Nussbeck, Lischetzke & Cole, 2010; Schermelleh-Engel & Schweizer, 2008), mit welchem sich Trait-Effekte, Methoden-Effekte und Messfehler statistisch separieren lassen (Eid et al., 2003). Das finale Modell umfasst vier *zielschülerspezifische Traits* (undiszipliniertes, dissoziales Verhalten; affektives, kognitives Störungsempfinden), welche mit allen drei "*Methoden*" (Teacher-, Peer-, Self-Ratings) gemessen wurden.

Trait-Effekte werden nach dem CT-C(M-1)-Ansatz basierend auf Trait-Faktoren anhand einer vorgängig bestimmten Referenzmethode geschätzt – im vorliegenden Fall wurden dafür die aggregierten Peer-Ratings herangezogen. Darüber hinaus enthalten die Trait-Faktoren diejenigen Anteile der Ratings aller anderen Methoden, welche mit den aggregierten Peer-Ratings konvergieren. Damit entsprechen die Trait-Faktoren einer *gruppenübergreifend intersubjektiv geteilten Sichtweise aller Rater*.

Bei den *Methoden-Effekten* handelt es sich um Rater-Effekte, wobei zwei Arten berücksichtigt werden. Bei der ersten Art handelt es sich um *gruppenspezifische Rater-Effekte*. Das sind mittlere Abweichungen der Teacher- und Self-Ratings von der gemeinsamen Sichtweise. Diese Abweichungen werden anhand der unstandardisierten Trait-Faktorladungen geschätzt: Sind die Ladungen grösser als 1.00, dann bedeutet dies eine durchschnittliche Überschätzung eines Ratertyps im Vergleich zur Referenzmethode. Sind die Ladungen kleiner als 1.00 entspricht dies einer durchschnittlichen Unterschätzung durch diesen Ratertyp. Es handelt sich folglich um systematische Abweichungen, welche auf die Zugehörigkeit zu einem bestimmten Ratertyp (Teacher-, Self-Rater) zurückzuführen sind.

Bei der zweiten Art handelt es sich um *personenspezifische Rater-Effekte*, die im Modell basierend auf *Methoden-Faktoren* geschätzt werden. Darin enthalten sind – unter Kontrolle des oben genannten gruppenspezifischen Rater-Effekts – weitere systematische Abweichungen einzelner Ratings im Vergleich zu den anderen Ratings desselben Ratertyps. Für die Referenzmethode wird nach dem CT-C(M-1)-Ansatz kein Methoden-Faktor modelliert – deswegen "M-1" (Eid et al., 2008). Im vorliegenden Fall wurde also für die aggregierten Peer-Ratings kein Methoden-Faktor geschätzt (jedoch für die unigen Peer Ratings).

Die SUGUS-Daten sind *hierarchisch (genestet) strukturiert*, sodass in Anlehnung an Carretero-Dios, Eid und Ruch (2011) und Eid et al. (2008) ein twolevel-CT-C(M-1) Modell erstellt wurde. Abbildung 14 illustriert dieses Modell inklusive der standardisierten Resultate (Ladungsparameter und Korrelationen) der Modellschätzung.

Ebene 2 (L2) ist die Ebene der Zielschüler/innen ($n=1'677$), wo die Self- und Teacher-Ratings sowie die aggregierten Peer-Ratings angesiedelt sind. Pro Zielschüler/in wurden durchschnittlich 3.47 Peer-Ratings eingeholt, welche auf L2 zu messfehlerbereinigten Random Intercepts (True Scores) latent aggregiert wurden. Die ursprünglichen, unigen Peer-Ratings sind hingegen auf Ebene 1 (L1) angesiedelt, wo die Fallzahl 5811 unique Ratings beträgt.

Bei den *Trait-Faktoren* (L2) handelt es sich um die mittlere Einschätzung der Peers pro Zielschüler/in sowie um denjenigen Anteil der Teacher- und Self-Ratings, welcher mit den mittleren Peer-Ratings übereinstimmt – es ist die Schnittmenge der verschiedenen Perspektiven auf die Zielschüler/innen. Bei den Devianz Ratings kann diese gruppenübergreifend intersubjektiv geteilte Sichtweise als Annäherung an den interessierenden *objektiven Kern* von Unterrichtsstörungen aufgefasst werden. Beim Störungsempfinden messen die Trait-Faktoren die von allen Ratern durchschnittlich empfundene Störungsintensität pro Zielschüler/in.

Die gruppenspezifischen Abweichungen der Teacher- und Self-Ratings von dieser gemeinsamen Sichtweise sind in den unstandardisierten Trait-Faktorladungen enthalten (diese sind dokumentiert in der Originalarbeit (Eckstein, under review)). Die personenspezifischen Rater-Effekte werden in den *Methoden-Faktoren* der Teacher-Ratings und der Self-Ratings (L2) sowie in den unigen Methoden-Faktoren der Peer-Ratings (L1) geschätzt.

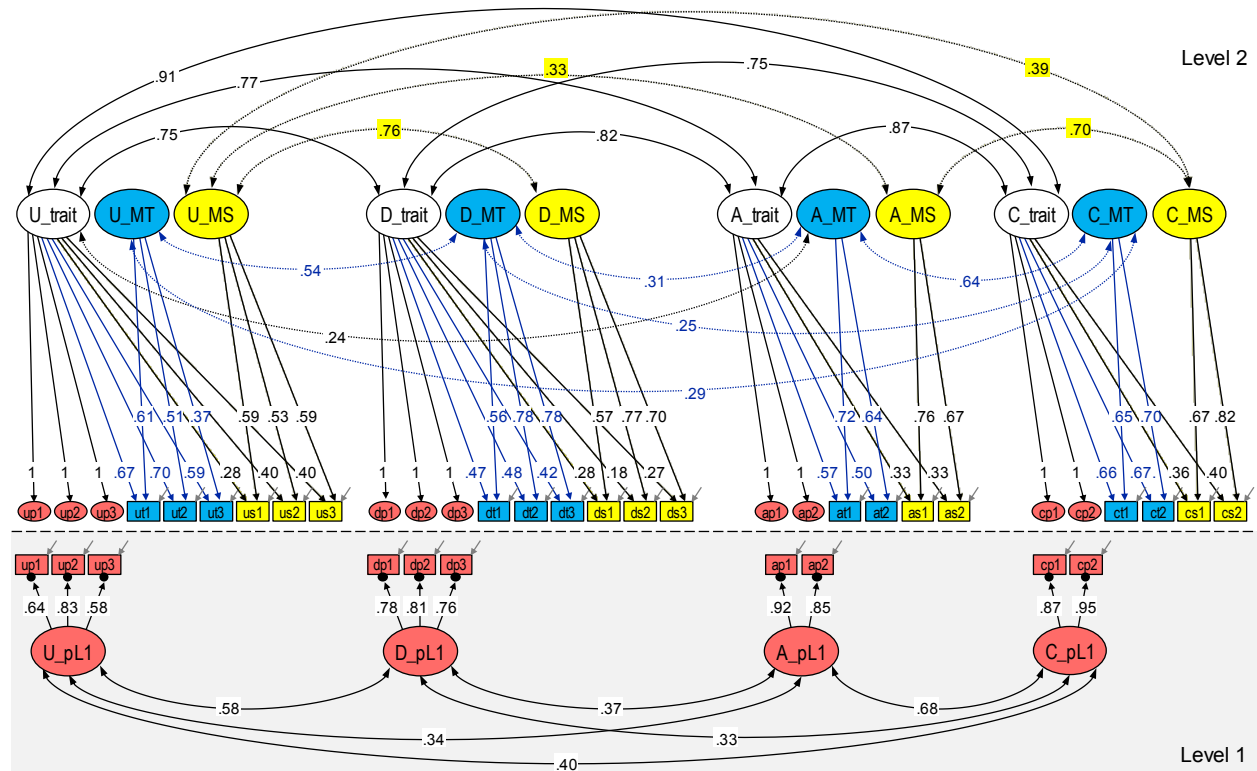


Abbildung 14. Resultate des Twolevel CT-C(M-1) Modells (Eckstein, under review).

Anmerkungen: Alle abgebildeten Parameter sind signifikant ($p < .05$), auf eine Kennzeichnung mit Asterisk (*) wurde verzichtet. Korrelationen sind nur mit einem Betrag $|r| > 0.20$ eingezeichnet. Ebenfalls nicht eingezeichnet sind die auf 0 restringierten Korrelationen zwischen Trait- und Methoden-Faktoren der gleichen Trait-Method-Einheit (diese sind per Definition unkorreliert (Eid et al., 2008)).

Auf Ebene 2 (L2) illustrieren die weissen Ovale die Trait-Faktoren (U_trait: undiszipliniertes Verhalten; D_trait: dissoziales Verhalten; A_trait: affektives Störungsempfinden; C_trait: kognitives Störungsempfinden). Die vier blauen Ovale auf L2 illustrieren die Methoden-Faktoren für die Teacher-Ratings (U_MT, D_MT, A_MT, C_MT); die vier gelben Ovale auf L2 illustrieren die Methoden-Faktoren der Self-Ratings (U_MS, D_MS, A_MS, C_MS).

Die zehn roten, kleinen Ovale auf L2 illustrieren die pro Zielschüler/in latent aggregierten Peer-Ratings (up1-up3, dp1-dp3, ap1-ap2, cp1-cp2). Es handelt sich um messfehlerbereinigte Random Intercepts. Die zehn blauen Kästchen stehen für die manifesten Indikatoren der Teacher-Ratings (ut1-ut3, dt1-dt3, at1-at2, ct1-ct2); die zehn gelben Kästchen illustrieren die manifesten Indikatoren der Self-Ratings (us1-us3, ds1-ds2, as1-as2, cs1-cs2). Die grauen Pfeilchen illustrieren die indikatorspezifischen Messfehler.

Auf Ebene 1 (L1) werden die Abweichungen einzelner Peer-Ratings vom aggregierten Score in unquen Methoden-Faktoren modelliert, die als vier rote Ovale dargestellt sind (U_pL1, D_pL1, A_pL1, C_pL1). Die zehn roten Kästchen auf L1 illustrieren die manifesten Indikatoren der unquen Peer-Ratings, welche auf L1 auch einen Messfehler aufweisen (graue Pfeilchen).

Das Modell passt trotz des infolge der hohen Fallzahl signifikanten χ^2 -Werts gut zu den Daten ($\chi^2[\text{MLR}] = 1360.18$, $df = 372$, $p < .001$; $\text{RMSEA} = .021$; $\text{CFI} = 0.93$; $\text{SRMR}_{L1} = 0.04$; $\text{SRMR}_{L2} = 0.04$). Basierend auf den Resultaten dieser Modellschätzung wurden nach den Formeln von Eid et al. (2008) nachträglich drei Varianzkomponenten der Teacher- und Self-Ratings sowie der unique Peer-Ratings pro Indikator sowie pro Konstrukt manuell berechnet. Eine vollständige Übersicht über alle Koeffizienten der drei Varianzkomponenten finden sich in der Originalarbeit (Eckstein, under review).

- (1) *Konsistenz*: Varianzanteil, welcher mit den aggregierten Peer-Ratings konvergiert (Ausmass der Übereinstimmung; Interrater Agreement).
- (2) *Spezifität*: Varianzanteil, welcher nicht mit den aggregierten Peer-Ratings konvergiert (Ausmass des Rater-Effekts).
- (3) *Unreliabilität*: Anteil unerklärter Varianz (Ausmass des Messfehlers).

Die Analysen ergaben *weitgehend erwartungskonforme Resultate*. Die Korrelationen zwischen den Trait-Faktoren (L2) zeigen, dass die Zielschüler/innen in der von allen Ratern geteilten Sichtweise als umso störender empfunden werden, je häufiger sie sich deviant verhalten. Doch diese gemeinsame Sichtweise fällt anteilmässig ziemlich bescheiden aus, wie die Varianzkomponenten zeigen: Über alle Konstrukte hinweg lässt sich die Varianz der Ratings lediglich zu 7 bis 54 Prozent durch Trait-Effekte erklären – demgegenüber sind 30 bis 70 Prozent der Varianz raterspezifisch und demnach auf Rater-Effekte zurückzuführen. Diese Befunde erhärten die Kernannahmen des SUGUS-Projekts: *Die Intensität des Störungsempfindens hängt zwar stark mit der Auftretenshäufigkeit normabweichenden Verhaltens zusammen, doch die Störungsrezeption gestaltet sich in vielerlei Hinsicht als höchst subjektiv.*

Darüber hinaus gab es auch unerwartete Resultate. Beispielsweise wurde angenommen, dass die *Auftretenshäufigkeit unterrichtlicher Devianz* wie beim Pretest (Kap. 2.2.9) weitgehend konsistent eingeschätzt wird²⁸. Bei der Hauptstudie zeigte sich aber, dass die verschiedenen Ratingtypen *maximal zu 54 Prozent* mit den mittleren Peer-Ratings (Referenzmethode) konvergieren. Demgegenüber erklären *Rater-Effekte bis zu 70 Prozent* der Varianz des eingeschätzten Schülerverhaltens.

Diese Resultate bewegen sich zwar in einem für die Unterrichtsforschung üblichen Bereich²⁹ – mit Blick auf die Teacher-Ratings der Skala undiszipliniertes Verhalten (54 Prozent Konsistenz) ist der geteilte Varianzanteil sogar als sehr hoch zu beurteilen. Insgesamt ist jedoch zu konstatieren, dass die Rater die Auftretenshäufigkeit devianten Verhaltens – trotz niedrig-inferenter Operationalisierung – unterschiedlicher eingeschätzt haben als angenommen.

Nachträglich lassen sich diese Divergenzen zwischen den Ratings wohl mit etikettierungstheoretischen (Kap. 2.2.4) oder mit wahrnehmungspsychologischen Argumenten (Kap. 2.2.5) erklären. Doch es bleibt festzuhalten, dass ein methodisches Ziel der Studie nicht vollständig erreicht wurde: Die aufwändige, niedrig-inferente Operationalisierung des Instruments (Kap. 2.2.8) reduzierte den Interpretationsspielraum für die Befragten weniger deutlich als erhofft. Dies zeigte sich auch in der nicht besonders deutlichen Übereinstimmung zwischen den multiplen Peer-Ratings pro Zielschüler/in hinsichtlich der Devianz-Inzidenz (18 – 31 % Konsistenz).

²⁸ Zur Erinnerung: Beim Pretest wurde die Konsistenz der Peer-Ratings mittels Intraklassenkorrelationen berechnet. Die bei der Hauptstudie berechneten Konsistenz-Koeffizienten stellen ein analoges Mass dar. Im Falle der Varianz der messfehlerbereinigten True Scores der Peer-Ratings sind die beiden Koeffizienten identisch (siehe Fussnote 13, S. 66).

²⁹ Beispielsweise konvergieren multiple Einschätzungen von Schülerinnen und Schülern zu Merkmalen des Unterrichts üblicherweise nur zu etwa 10–20% (Wagner et al., 2016).

In Bezug auf das *subjektive Störungsempfinden* wurden hohe raterspezifische Anteile erwartet – und auch ermittelt (14 – 47% Konsistenz vs. 48 – 70% Spezifität). Einzig beim *kognitiven Störungsempfinden* der Teacher-Ratings fiel das Verhältnis weniger deutlich aus als angenommen (47% Konsistenz vs. 48% Spezifität). Dieses Resultat überrascht ein wenig, weil es sich um ein raterspezifisches Mass handelt und daher ein noch höherer Anteil an Spezifität erwartet wurde. Aus theoretischen Überlegungen wurde angenommen, dass es weitgehend subjektiv ist, inwieweit die Rater die Zielschüler/innen als störend/ablenkend empfinden. Die Analysen haben aber gezeigt, dass die Lehrpersonen und die Peers etwa zur Hälfte darin übereinstimmen, wie stark einzelne Zielschüler/innen im Unterricht für Ablenkung sorgen und ihre Konzentration beeinträchtigen. Interessanter Weise schätzten die Peer-Rater den kognitiven Störungsgrad der Zielschüler/innen untereinander weniger übereinstimmend ein (23% Konsistenz).

Beim *affektiven Störungsempfinden* ist das Verhältnis bei allen Ratingtypen erwartungskonform: Die Spezifität überwiegt mit 53 – 67 Prozent Varianzanteil deutlich (gegenüber 14 – 33 Prozent Konsistenz). Damit wurde ein weiteres Resultat des Pretests (Kap. 2.2.9) repliziert: Zwar stimmen Lehrpersonen und Schüler/innen vergleichsweise genau darin überein, inwieweit sie Zielschüler/innen als ablenkend oder konzentrationsstörend beurteilen – jedoch ist es aber viel stärker subjektiv geprägt, inwieweit sie sich darüber hinaus auch über diese Zielschüler/innen ärgern oder sie als unfreundlich beurteilen.

Die Konsistenz-Koeffizienten der Teacher-Ratings fallen bei sämtlichen Konstrukten höher aus als diejenigen der unigen Peer-Ratings. Demnach *konvergieren die Einschätzungen der Lehrpersonen stärker* mit der Referenzmethode, als diese mit den unigen Peer-Ratings, aus denen sie ursprünglich berechnet wurde. Das könnte darauf zurückzuführen sein, dass einige Schüler/innen bei der Befragung zu Übertreibungen neigten, während andere eher untertrieben haben. Solche mutmasslichen interpersonellen Schwankungen dürften sich bei der Aggregation auf der Ebene der Zielschüler/innen gegenseitig ausgemittelt haben, sodass die aggregierten Scores näher an den (vermutlich weniger extremen) Einschätzungen der Lehrperson liegen als an den einzelnen Peer-Ratings.

Die *Korrelationen zwischen den Trait-Faktoren (L2)* erlauben Rückschlüsse auf die Beziehungen zwischen den Konstrukten in der von allen Ratern geteilten Sichtweise: Zielschüler/innen, welche sich häufiger undiszipliniert verhalten, zeigen auch vermehrt dissoziales Verhalten ($r = .75^*$). Darüber hinaus werden Zielschüler/innen, welche sich häufiger deviant verhalten, als stärker störend wahrgenommen ($.77^* \leq r \leq .91^*$). Diese Befunde wurden zwar erwartet, doch sind die Koeffizienten wie beim Pretest derart hoch (Kap. 2.2.9), dass sie kritisch zu diskutieren sind. Insbesondere die Korrelation zwischen undiszipliniertem Verhalten und kognitivem Störungsempfinden ($r = .91^*$).

Bei der Nachbereitung der Vorstudie wurde gemutmasst, dass die Vorselektion der Zielschüler/innen durch die Lehrpersonen eine Überzeichnung des Kontrasts zwischen den Zielschülerinnen und Zielschülern begünstigt haben könnte, sodass die extreme Höhe des Koeffizienten womöglich ein Artefakt darstelle (Kap. 2.2.11). Die Replikation des Ergebnisses mit einer neuen, mit geringerem Bias verbundenen Samplingstrategie (Kap. 2.3.2) legt nun aber

einen alternativen Erklärungsversuch nahe: Vermutlich haben die Rater die Items zum kognitiven Störungsempfinden als ereignisnah interpretiert (z.B. "...hat mich im Unterricht abgelenkt"). Dies dürfte zur Folge gehabt haben, dass die intersubjektiv geteilte Beurteilung der Störungsintensität stark mit den tatsächlichen Ereignissen korreliert, welche anhand der geteilten Devianz-Ratings geschätzt wurden (z.B. "Hat im Unterricht mit einem anderen Kind geredet, obwohl es ruhig sein sollte").

In der Konsequenz bedeutet dies, dass sich die theoretisch entflochtenen Merkmalsdimensionen gestörten Unterrichts (Kap. 2.2.2 – 2.2.6) in dieser geteilten Sichtweise nicht klar unterscheiden lassen. Aus methodischer Sicht ist jedoch hinzuzufügen, dass die Konstrukte nur unter Ausschluss der Rater-Effekte derart hoch korrelieren – also *nur bei exklusiver Betrachtung der Schnittmenge* an geteilter Varianz der drei Ratingtypen. Und diese Schnittmenge fällt anteilmässig ziemlich bescheiden aus. Der Löwenanteil der Störungsrezeption geht auf subjektive, raterspezifische Anteile zurück.

Unter Berücksichtigung der Rater-Effekte ist demnach davon auszugehen, dass sich die Merkmalsdimensionen sehr wohl unterscheiden lassen. Das zeigt sich (auch) darin, dass die Methoden-Faktoren weit weniger hoch korrelieren als die Trait-Faktoren. Darüber hinaus unterscheiden sich die Konstrukte nach theoretischen Gesichtspunkten derart klar voneinander, dass nach Kleinke, Schlüter und Christ (2017) davon abgesehen wurde, die Trait-Faktoren in der Modellschätzung zusammenzufassen. Nicht zuletzt ist die Aufrechterhaltung der getrennten Modellierung der zwei Konstrukte aber auch der weiterführenden Analysestrategie geschuldet, deren Ziel darin besteht, Bedingungen von Devianz einerseits und vom Störungsempfinden andererseits zu untersuchen (Kap. 3.4).

Die *Korrelationen zwischen den Methoden-Faktoren* deuten darauf hin, dass sich die personenspezifischen Rater-Effekte weitgehend generalisieren lassen: Rater, welche eine/n Zielschüler/in bei einem Konstrukt überschätzen (verglichen mit der durchschnittlichen Einschätzung ihres Ratertyps), überschätzen diese/n Zielschüler/in tendenziell auch bei den anderen Konstrukten. Oder vice versa: Unterschätzung bei einem Konstrukt geht mit Unterschätzung bei den anderen Konstrukten einher. Diesbezüglich ist der Vollständigkeit halber noch zu erwähnen, dass die Ursachen dieser generellen Rater-Effekte noch nicht untersucht worden sind. In der bisherigen theoretischen Argumentation wurden vorrangig Merkmale der Rater als Bedingungen ihrer Über- bzw. Unterschätzung thematisiert. Es könnte aber auch sein, dass Merkmale der Zielschüler/innen die Rater-Effekte verursachen, etwa wenn sie sich tatsächlich öfters deviant verhalten im Vergleich zum "Durchschnittsschüler".

Schliesslich wurden *Korrelationen zwischen Trait-Faktoren und Methoden-Faktoren* ermittelt, wobei fast alle Koeffizienten einen Betrag von $|r| < 0.20$ aufweisen und/oder nicht signifikant sind. Die grösste Korrelation besteht zwischen U_trait und A_MT ($r = .24^*$) und deutet folgende Tendenz an: Lehrpersonen ärgern sich mehr über Zielschüler/innen – verglichen mit dem durchschnittlichen Teacher-Rating (A_MT), wenn sich die Zielschüler/innen in der von allen Ratern geteilten Sichtweise häufiger undiszipliniert verhielten (U_trait). Diese Korrelation ist aber als schwacher Effekt zu beurteilen (Cohen, 1988).

3. Zusammenfassung und Bilanz

Die vorliegende, kumulativ angelegte Dissertation entstand vor dem Hintergrund der bildungspolitisch angestossenen Entwicklung in Richtung einer integrativen Volksschule. Dieser Trend wird in der (Fach-)Öffentlichkeit kontrovers diskutiert (Trautmann & Wischer, 2011). Dabei stehen weniger die Ideale der Integration in der Kritik, als dass Schwierigkeiten bei der Umsetzung moniert werden (Reusser et al., 2013). Beispielsweise fragen sich viele Lehrpersonen, wie sie Schüler/innen mit Verhaltensauffälligkeiten erfolgreich integrieren können (Bernhardt, 2007), sodass diese den Unterricht nicht zu sehr stören. In der vorliegenden Dissertation wurde argumentiert, dass diese individualpsychologische Fokussierung auf störende Schüler/innen das Phänomen gestörten Unterrichts einseitig und unvollständig beleuchtet. Es wurde ein interaktionistischer, stärker ganzheitlicher Theorierahmen entwickelt, wonach Unterrichtsstörungen aus einem Element der Produktion (stören) und aus einem Element der Rezeption (gestört sein) bestehen. *Diese beiden Elemente gestörten Unterrichts wurden im Rahmen von vier Originalarbeiten empirisch untersucht.* Die Ergebnisse der vier Beiträge sowie weiterführende Überlegungen wurden im Manteltext der Dissertation anhand einer theoretischen Einordnung zu einem Gesamtbild integriert.

In diesem Kapitel 3 werden die *Haupterträge der Dissertation* dargelegt und diskutiert. Dazu werde ich in den nachfolgenden Kapiteln 3.1.1 bis 3.1.4 die Entstehungskontexte, die Ziele und die wichtigsten Ergebnisse der vier Originalarbeiten sowie Anschlussüberlegungen jeweils in Form einer Auflistung nochmals komprimiert darstellen. Hernach leite ich in Kap. 3.2 übergeordnete Erträge aus dem Gesamtbild der vier Beiträge ab. In Kapitel 3.3 thematisiere ich Limitationen der unternommenen Analysen und in Kapitel 3.4 gebe ich einen Ausblick auf mögliche weiterführende Analysen zur Untersuchung von Ursachen gestörten Unterrichts.

3.1. Haupterträge der einzelnen Originalarbeiten

3.1.1. Beitrag I: Schüler/innen mit Verhaltensauffälligkeiten in der Regelschule

- *Entstehungskontext:* Heterogenitäts-Projekt (standardisierte Online-Befragung von 185 Klassenlehrpersonen)
- *Zielsetzung:* Analyse von Bedingungen optimistischer Überzeugungen von Lehrpersonen hinsichtlich der Realisierbarkeit der Integration mittels eines Strukturgleichungsmodells
- *Hauptergebnis 1 – positiver Einfluss einer erweiterten didaktischen Unterrichtspraxis:* Lehrpersonen, die angeben, regelmässig binnendifferenziert und auf der Grundlage erweiterter Lehr-Lernformen unterrichten, sind vergleichsweise optimistisch im Hinblick auf die Realisierbarkeit der Integration ($\beta_{\text{Primar}} = .50^*$; $\beta_{\text{Sek}} = .53^*$).
- *Hauptergebnis 2 – negativer Einfluss der Quote von Schülerinnen und Schülern mit Verhaltensauffälligkeit:* Je mehr Schüler/innen mit Verhaltensauffälligkeiten anteilmässig in einer Klasse sind, desto weniger optimistisch ist die Lehrperson ($\beta_{\text{Primar}} = -.12^*$; $\beta_{\text{Sek}} = -.25^*$).

Oder umgekehrt: Je skeptischer eine Lehrperson gegenüber der Integration eingestellt ist, desto eher erachtet sie bestimmte Verhaltensweisen als unangemessen und etikettiert entsprechend mehr Schüler/innen als verhaltensauffällig.
- *Anschlussüberlegung:* Die eigentliche Herausforderung bei Verhaltensauffälligkeiten in der Regelschule sind nicht Individualmerkmale einzelner Schüler/innen (z.B. ADHS), sondern Störungen des Unterrichts, die daraus hervorgehen, dass die beteiligten Akteure durch normabweichendes Schülerverhalten gestört werden.

3.1.2. Beitrag II: Gestörter Unterricht – Entflechtung zentraler Merkmalsdimensionen

- *Entstehungskontext:* SUGUS-Projekt; quantitative Teilstudie; Pretest
(Fragebogenerhebung in 5. Klassen; 11 Lehrpersonen; 239 Schüler/innen)
- *Zielsetzung 1:* Darstellung des interaktionistischen Theorierahmens des SUGUS-Projekts
(Bedingungen der Produktion und Rezeption gestörten Unterrichts)
- *Zielsetzung 2:* Qualitätsprüfung der Instrumente zur Erfassung (a) der Auftretenshäufigkeit unterrichtlicher Devianz einzelner Zielschüler/innen sowie (b) des zielschülerspezifischen, subjektiven Störungsempfindens der Lehrpersonen und Schüler/innen
- *Hauptergebnis 1 – ICC-Koeffizienten einer twolevel CFA:* Bei der Devianz-Inzidenz stimmen die Ratings mehrerer Mitschüler/innen (Peers) pro Zielschüler/in mit 71% stark überein, während sie beim affektiven Störungsempfinden mit 24 % weniger übereinstimmen.
- *Hauptergebnis 2 – ebenenspezifische Korrelationen der twolevel CFA:* In den Perspektiven der einzelnen Peers (L1) korreliert das subjektive Störungsempfinden weniger stark mit dem normabweichenden Verhalten der von ihnen eingeschätzten Zielschüler/innen ($.08 \text{ n.s.} \leq r \leq .46^*$) als in der von allen Peers intersubjektiv geteilten Sichtweise (L2; $.84^* \leq r \leq .97^*$).
- *Hauptergebnis 3 – Problemausmass:* Die Zielschüler/innen haben im Mittel nur selten deviantes Verhalten gezeigt; die Befragten empfanden sie im Durchschnitt als wenig störend.
- *Zusatzertrag:* Die Resultate klassenspezifischer Analysen wurden den teilnehmenden Lehrpersonen in Form personalisierter Ergebnisberichte zurückgemeldet.

3.1.3. Beitrag III: Deskriptiver Perspektivenvergleich

- *Entstehungskontext:* SUGUS-Projekt; quantitative Teilstudie; Hauptstudie
(Fragebogenerhebung in 5. Klassen; 85 Lehrpersonen; 1'412 Schüler/innen)
- *Zielsetzung 1:* Untersuchung des Problemausmasses in Bezug auf vier Kategorien: Devianz-Inzidenz und subjektives Störungsempfinden (beides zielschülerspezifisch); allgemeine Störungssensitivität und allgemeines Belastungserleben (beides raterspezifisch)
- *Zielsetzung 2:* Perspektivenvergleiche auf der Ebene der Stichprobe und einzelner Klassen
- *Hauptergebnis 1 – Problemausmass:* Aus Sicht der Lehrpersonen zeigten die Schüler/innen der Durchschnittsklasse gemeinsam im Mittel 4.59 (meist bagatellhafte) Normabweichungen pro Lektion; 91.9 % aller Schüler/innen wurden von ihren Lehrpersonen als kaum störend beschrieben. Das mittlere Problemausmass ist gering (analog Hauptergebnis 3 von Beitrag II); es wurden aber auch einzelne Klassen mit hohem Problemausmass identifiziert.
- *Hauptergebnis 2 – Perspektivendifferenzen:* Auf Stichprobenebene wurden nur geringfügige Mittelwertsdifferenzen gefunden. Jedoch konnten innerhalb einzelner Schulklassen teils markante Unterschiede zwischen der Perspektive der Lehrperson und derjenigen ihrer Schüler/innen ermittelt werden.
- *Zusatzertrag:* Wie beim Pretest wurden die Resultate der klassenspezifischen Analysen den teilnehmenden Lehrpersonen in Form personalisierter Ergebnisberichte zurückgemeldet. Die in den Berichten dargestellten (teils markanten) Perspektivendifferenzen wurden von einigen Lehrpersonen mit Erstaunen zur Kenntnis genommen, wie im Rahmen der qualitativen SUGUS-Teilstudie festgestellt wurde.

3.1.4. Beitrag IV: Konsistenz und Spezifität der verschiedenen Perspektiven

- *Entstehungskontext:* quantitative SUGUS-Hauptstudie (analog Beitrag III)
- *Vorbereitende Analysen:* Skalierung der Konstrukte "Devianz" und "Störungsempfinden" in je zwei Dimensionen: undiszipliniertes, dissoziales Verhalten; affektives, kognitives Störungsempfinden (siehe Technischer Bericht (Eckstein et al., 2018))
- *Zielsetzung 1:* Ermittlung der Konsistenz und Spezifität der Teacher-, Peer-, Self-Ratings
- *Zielsetzung 2:* Analyse der Zusammenhänge von Devianz und Störungsempfinden in einer von allen Ratern geteilten Sichtweise
- *Zielsetzung 3:* Untersuchung der Generalisierbarkeit von Rater-Effekten
- *Methode – simultane Analyse der drei Ratingtypen (Teacher-, Peer-, Self-Ratings) mittels eines twolevel CT-C(M-1) Modells.* Dieses Verfahren schätzt die Schnittmenge an konsistenten Einschätzungen aller Rater (Trait-Effekte) sowie raterspezifische Abweichungen von dieser gemeinsamen Sichtweise (Rater-Effekte). Dazu wurden die multiplen Peer-Ratings pro Zielschüler/in aggregiert und als Referenzmethode herangezogen. Sie bilden den Vergleichsstandard, womit das Ausmass der Rater-Effekte bestimmt wurde.
- *Hauptergebnis 1 – Konsistenz und Spezifität in Bezug auf die Auftretenshäufigkeit undisziplinierten Verhaltens (z.B. schwatzen):* Die Teacher-Ratings stimmen zu 54% mit den mittleren Peer-Ratings überein, 30% der Varianz der Teacher-Ratings sind spezifisch. Bei diesem unerwartet hohen Rater-Effekt handelt es sich im Durchschnitt eine Überschätzung durch die Lehrpersonen – oder um eine Unterschätzung durch die Peers.

- *Hauptergebnis 2 – Konsistenz und Spezifität in Bezug auf die Auftretenshäufigkeit dissozialen Verhaltens (z.B. andere Kinder beleidigen):* 26% der Varianz der Teacher-Ratings lassen sich mit den mittleren Peer-Ratings erklären. Der raterspezifische Anteil beläuft sich auf 61%. Dieser unerwartet hohe Rater-Effekt bedeutet eine durchschnittliche Unterschätzung aufseiten der Lehrperson – oder eine Überschätzung durch die Peers. Also gerade das Gegenteil vom zuvor beschriebenen Rater-Effekt bezüglich Undiszipliniertheit.
- *Hauptergebnis 3 – Konsistenz und Spezifität in Bezug auf das affektive Störungsempfinden (z.B. sich ärgern).* Die Einschätzungen der Lehrpersonen konvergieren zu 33% mit den mittleren Peer-Ratings, die Spezifität beläuft sich auf 53%. Verglichen mit den Peers haben die Lehrpersonen die Zielschüler/innen im Durchschnitt als weniger ärgerlich beschrieben.
- *Hauptergebnis 4 – Konsistenz und Spezifität in Bezug auf das kognitive Störungsempfinden (z.B. abgelenkt sein).* Die Einschätzungen der Lehrpersonen konvergieren zu 47% mit den aggregierten Peer-Ratings. Im Gegensatz zu den klar subjektiv geprägten Affekten sind sich Lehrpersonen und Schüler/innen also eher darin einig, welche Kinder sie im Unterricht ablenken oder in ihrer Konzentration beeinträchtigen.
- *Hauptergebnis 5 – geringe Konsistenz der Self-Ratings:* Die Selbstwahrnehmung der Zielschüler/innen konvergiert lediglich zu 7–20% mit den mittleren Peer-Ratings. Dagegen lassen sich 51-70% der Varianz mit Rater-Effekten erklären. Dabei handelt es sich jeweils um Unterschätzungen, die sich vermutlich mit einem selbstwertdienlichem Antwortverhalten der Zielschüler/innen erklären lassen.

- *Schlussfolgerung aus den Hauptergebnissen 1 – 5:* Rater-Effekte erklären 48–70% Varianz des Störungsempfindens der drei Ratertypen; bei den Devianz Skalen sind es 30–70%. Diese Ergebnisse zeigen, dass sich die Rezeption gestörten Unterrichts grösstenteils durch subjektive, raterspezifische Anteile erklären lässt. Als Ursachen dieser teilweise beträchtlichen Perspektivendifferenzen werden rollenspezifische und individuelle Merkmale der wahrnehmenden Akteure sowie kontextuelle Bedingungen und differenzielle Opportunitäten der Sichtbarkeit devianten Verhaltens angenommen.
- *Hauptergebnis 6 – hohe Korrelationen zwischen Devianz und Störungsempfinden in der von allen Ratern geteilten Sichtweise:* Die Trait-Faktoren sind hoch korreliert ($.77^* \leq r \leq .91^*$). Je häufiger sich die Zielschüler/innen gemäss intersubjektiv konsistenter Einschätzung normabweichend verhalten, desto höher ist die von ihnen ausgelöste Störungsintensität bei allen Ratern. Allerdings korrelieren die Konstrukte nur bei exklusiver Betrachtung der (kleinen) Schnittmenge an geteilter Varianz der drei Ratingtypen derart hoch.
- *Hauptergebnis 7 – niedrigere Korrelationen zwischen Devianz und Störungsempfinden in den raterspezifischen Perspektiven:* Die Methoden-Faktoren korrelieren weniger stark miteinander ($.21 \text{ n.s.} \leq r \leq .76^*$) als die Trait-Faktoren. Das heisst, das subjektive Störungsempfinden korreliert weniger stark mit Devianz in den raterspezifischen Perspektiven als in der geteilten Sichtweise (analog Hauptergebnis 2 von Beitrag II).
- *Hauptergebnis 8 – Generalisierbarkeit der Rater-Effekte:* Individuelle Über- oder Unterschätzung bei einem Konstrukt geht tendenziell auch mit Über- oder Unterschätzung bei den anderen Konstrukten einher.

3.2. Übergeordnete Erträge der vier Originalarbeiten

In diesem Kapitel 3.2 lege ich die übergeordneten Erträge dar, die sich aus den vier Originalarbeiten ableiten lassen. In Kapitel 3.2.1 thematisiere ich methodologische Implikationen für die wissenschaftliche Erforschung von Unterrichtsstörungen. Anschliessend behandle ich Implikationen für die Praxis (Kap. 3.2.2). Eine Darstellung ethischer Probleme in Bezug auf die diagnosebasierte Etikettierung von "Problemschülern" schliesst das Kapitel ab.

3.2.1. Methodologische Implikationen für die Erforschung von Unterrichtsstörungen

Wie die deskriptiven Analysen gezeigt haben, sind die zentralen Merkmalsdimensionen von *Unterrichtsstörungen stark rechtsschief* verteilt: Die meisten Schüler/innen verhalten sich nur selten normabweichend und werden von ihren Lehrpersonen, Mitschülerinnen und Mitschülern entsprechend als wenig störend empfunden. Diese Befunde decken sich weitgehend mit früherer Forschung (Müller & Hofmann, 2016), stehen aber im Widerspruch zum medial verbreiteten Bild, die (integrative) Schule habe flächendeckend mit massiven Störungen zu kämpfen (Munn et al., 2013). Zwar wurden auch in der SUGUS-Studie Klassen mit einem hohen Problemausmass identifiziert, doch handelt es sich hierbei um wenige Extremfälle.

Die Übereinstimmung der SUGUS-Befunde mit früheren Forschungsergebnissen (Müller & Hofmann, 2016) bezüglich der rechtsschiefen Verteilung gibt Anlass zur Annahme, dass Unterrichtsstörungen generell *kein normalverteiltes Phänomen* darstellen. Diese Schlussfolgerung zieht drei forschungsmethodische Implikationen nach sich: (i) Operationalisierung des Antwortformats von Messinstrumenten; (ii) Operationalisierung von Messinstrumenten bezüglich des Inferenzgrads; (iii) Angemessenheit verschiedener statistischer Verfahren der Datenanalyse.

(i) *Operationalisierung des Antwortformats von Messinstrumenten.* Es gibt Instrumente, deren Antwortskala entgegen der rechtsschiefen Verteilung des Phänomens eine *Normalverteilung insinuieren*. Dies könnte dazu führen, dass das Problemausmass unpräzise erfasst wird. Als Beispiel mag die einschlägige "Teacher-Child Rating Scale" (T-CRS) herhalten, mit welcher u.a. die Subskala "externalisierendes Problemverhalten" anhand eines fünfstufigen Antwortformats eingeschätzt wird: 1 = not a problem, 2 = very mild problem, 3 = moderate problem, 4 = serious problem, 5 = very serious problem (Hightower et al., 1986). Mit dieser Anordnung der Antwortkategorien wird implizit unterstellt, dass es sich bei unproblematischen (1) wie auch bei hoch problematischen (5) Fällen um "Extremfälle" handelt. Bei rechtsschiefer Verteilung sind aber die unproblematischen Fälle (1) gerade keine Extremfälle, sondern der Normalfall.

Dieses Format könnte Lehrpersonen dahingehend beeinflussen, das durchschnittliche Problemverhalten ungefähr in der Mitte der Antwortskala oder knapp darunter als "mildes" (2) oder "moderates" (3) Problem zu verorten. Entsprechend fand beispielsweise Granot (2014) mit diesem Instrument einen Mittelwert von $M = 2.2$ für externalisierendes Problemverhalten von Jungen ($SD = 1.1$; $n = 50$). Zum Vergleich: die durchschnittlichen SUGUS Teacher-Ratings von deviantem Verhalten von Jungen sind mit $M = 1.07$ ($SD = 1.55$; $n = 801$) weniger hoch – trotz des gegen oben unbegrenzten Antwortformats. Dieser Vergleich ist inhaltlich zwar schwer zu deuten, deutet aber einen *Methoden-Effekt* an: Mit der T-CRS haben die Lehrpersonen durchschnittlich höhere Werte angegeben als mit dem SUGUS-Instrument. Dieser Effekt lässt sich wahrscheinlich damit erklären, dass die Lehrpersonen bei SUGUS eine konkrete Anzahl Ereignisse pro definierten Zeitraum eingeschätzt haben, während die ordinalskalierte T-CRS eine eher subjektive Bewertungen des Problemausmasses erfasst.

(ii) *Operationalisierung von Messinstrumenten bezüglich des Inferenzgrads.* Ein zweites Problem von Instrumenten wie der T-CRS stellt der hohe Inferenzgrad dar. Beispielsweise verlangt Item (1) "Disruptive in class" (Hightower et al., 1986) eine weitgehende Interpretationsleistung von den Lehrpersonen. Denn wie die SUGUS Analysen gezeigt haben, ist es hochgradig raterspezifisch, ob und inwieweit Schüler/innen als "störend" erachtet werden. Demzufolge wäre es wichtig, den *Interpretationsspielraum* bei Befragungen zu minimieren – sofern das Instrument das objektive Problemausmass und nicht einen subjektiven Eindruck der Befragten messen soll. Oder aber man thematisiert die Subjektivität solcher Beurteilungen explizit, indem man verschiedene Perspektiven einander gegenüberstellt.

(iii) *Angemessenheit verschiedener statistischer Verfahren der Datenanalyse.* Die rechtsschief verteilte Problemlage bei Unterrichtsstörungen hat neben der Thematik der adäquaten Datenerfassung auch Implikationen für die Datenauswertung. Diese zwei Aspekte sind miteinander verknüpft, denn wenn ein mutmasslich inadäquates Instrument normalverteilte Daten "generiert", können herkömmliche statistische Verfahren zum Einsatz gelangen, etwa parametrische Analysen (z.B. t-Tests, Regressionsanalysen, ANOVAs). Diese Verfahren sind bei schiefen Verteilungen im Prinzip aber nicht zulässig, auch wenn sie mit Einschränkungen und unter bestimmten Vorsichtsmassnahmen trotzdem anwendbar sind. Die linkssteil-rechtsschiefe Verteilungsform entspricht einer *negativen Binomialverteilung* (Hilbe, 2011, 2014). Studien zu Unterrichtsstörungen, welche diese Verteilungsform replizieren, können folgende Verfahren einsetzen:

- (1) Nichtparametrische Analysen (Rasch, Frieze, Hofmann & Naumann, 2010)
- (2) Berücksichtigung der negativen Binomialverteilung anhand der Poisson Modellierung (Coxe, West & Aiken, 2009)
- (3) Verteilungsfreie Schätzverfahren zur Analyse ordinalskalierter Daten, z.B. mit dem in Mplus implementierten WLSMV-Schätzer (weighted least squares mean- and variance-adjusted estimator) (Finney & DiStefano, 2013)
- (4) Maximum-Likelihood Schätzverfahren mit korrigierter χ^2 -Statistik und robusten Standardfehlern, z.B. mit dem in Mplus implementierten MLR-Schätzer (robust maximum likelihood) (Byrne, 2012)

Die aufgelisteten Optionen haben *unterschiedliche Vor- und Nachteile*, auf die an dieser Stelle nur oberflächlich eingegangen werden kann. Am passendsten wäre eigentlich die Poisson Modellierung (2), doch hat sich dieses Verfahren im Rahmen der SUGUS-Studie als nicht praktikabel erwiesen. Ein erstes Problem bereitet der immense Rechenaufwand – probenhalber initiierte SUGUS Analysen haben z.T. länger als 20 Stunden gedauert. Zudem sind in Mplus bei der Spezifikation "COUNT" zur Schätzung von Poisson Modellen (Muthén & Muthén, 2017) verschiedene Optionen nicht verfügbar, z.B. können nicht alle Parameter (z.B. Faktorladungen, Korrelationen) standardisiert werden, was die Interpretation der Resultate massiv erschwert. Deshalb beruhen die wichtigsten SUGUS Analysen der vorliegenden Arbeit auf den zwei letztgenannten Verfahren mit dem WLSMV-Schätzer (3) und dem MLR-Schätzer (4).

3.2.2. Implikationen für die Praxis

Für die Praxis sind die vorliegenden Ergebnisse insofern bedeutsam, als dass sie die Subjektivität der Störungsrezeption verdeutlichen. Es hat sich gezeigt, dass Unterrichtsstörungen *keine objektiven Tatsachen* sind, welche sich ausschliesslich auf die schülerseitige Produktion zurückführen lassen. Vielmehr handelt es sich um interaktional ko-konstruierte Phänomene. Dieses Verständnis ist in der Praxis nicht allgegenwärtig anzutreffen (Bernhardt, 2007). Demzufolge könnte es hilfreich sein, den interaktionistischen Problemcharakter gestörten Unterrichts im Rahmen der *Aus- und Fortbildung von Lehrpersonen* in einer reichhaltigen Sprache zu vermitteln. Dazu bedarf es geeigneter hochschuldidaktischer Tools.

Im Rahmen der Dissertation (Beiträge II und III) wurde ein solches Tool entwickelt: Den an der Studie teilnehmenden Lehrpersonen wurden Resultate klassenspezifischer Analysen in Form von *Netzdiagrammen* in personalisierten Berichten zugestellt. In diesen Diagrammen wurden jeweils die Perspektive der Lehrperson sowie diejenige ihrer Schüler/innen simultan dargestellt, wobei zum Teil beträchtliche Unterschiede in der Beschreibung des Problemausmasses sichtbar wurden. Einige Lehrpersonen haben diese Divergenzen mit Erstaunen zur Kenntnis, was gegenwärtig im Rahmen der qualitativen Teilstudie des SUGUS-Projekts vertiefend analysiert wird (Marusic-Würscher & Reusser, 2018).

Angesichts dieses Erstaunens scheint es gelungen zu sein, den Lehrpersonen die Unterschiedlichkeit der verschiedenen Sichtweisen anhand der in den Netzdiagrammen dargestellten Profile zu veranschaulichen. Demzufolge könnten *Visualisierungen* ähnlicher Art in unterstützender Funktion eingesetzt werden, um den Aspekt der differenziellen Wahrnehmung, Deutung und Bewertung gestörten Unterrichts durch die beteiligten Akteure in der Aus- und Fortbildung von Lehrpersonen zu besprechen und zu vermitteln.

Die Einsicht, dass es sich bei Unterrichtsstörungen um ein facettenreiches, konstruiertes Phänomen handelt, könnte als *Anstoss* dienen, dass Lehrpersonen ihre eigene subjektive Sichtweise hinterfragen und vermehrt versuchen, auch die Perspektive ihrer Schüler/innen einzunehmen. Festgestellte Gemeinsamkeiten und Divergenzen in der Störungswahrnehmung dürften sich auch im Klassenrat ergiebig besprechen lassen. Im besten Fall leiten Lehrpersonen ausgehend von solchen Prozessen pädagogisch-didaktische Ideen zum Umgang mit Störungen ab (Wettstein, 2013).

3.2.3. Ethische Implikationen in Bezug auf die Diagnostik von "Problemschülern"

Schliesslich implizieren die Befunde zur Subjektivität der Störungswahrnehmung ein *ethisches Problem* betreffend das diagnosebasierte Labeling einzelner "Problemschüler" (Hunt et al., 1989). Es wäre beispielsweise zu prüfen, inwieweit die Testinstrumente schulpsychologischer Abklärungen raterspezifische Anteile enthalten. Eine solche Prüfung würde Aufschluss darüber geben, inwieweit eine gestellte Diagnose auf das diagnostizierte Kind bzw. auf die diagnostizierende Person zurückzuführen ist. Korsch und Petermann (2012) haben beispielsweise darauf hingewiesen, dass unterschiedliche Quellen, welche zur Früherkennung von "Risikokindern" einbezogen werden (z.B. die Eltern), in ihren Einschätzungen teils weit auseinander liegen. Dabei handelt es sich nicht um ein Problem des akademischen Diskurses. Denn solche Diagnosen werden konkret zur *Begründung schulischer Laufbahnentscheidungen* herangezogen, z.B. für den Beschluss einer separativen Sonderschulung (Bildungsdirektion Kanton Zürich, 2013). Die Ergebnisse der vorliegenden Arbeit zeigen, dass äusserst sorgfältig abgewogen werden sollte, bevor es heisst: "Störenfriede raus!" (Burri, 2017).

3.3. Limitationen der Arbeit

Die Erträge der Dissertation sind mit Blick auf *vier Problemzonen* kritisch zu diskutieren: (i) Rater-Effekte trotz niedrig-inferenter Operationalisierung; (ii) Qualität der Stichprobe; (ii) Spannungsfeld Theorie – Empirie; (iv) bislang unerfüllte Ertragserwartungen. Im Folgenden werde ich diese Limitationen in vier Abschnitten darlegen.

(i) Rater-Effekte trotz niedrig-inferenter Operationalisierung. Das Forschungsdesign und die Auswertungsstrategie der quantitativen SUGUS Studie waren darauf ausgelegt, den *objektiven Kern* gestörten Unterrichts approximativ zu bestimmen. Im Rahmen der CT-C(M-1) Modellierung ist das zwar grundsätzlich gelungen, indem eine gemeinsame Sichtweise mehrerer Informanten unter Berücksichtigung von Rater-Effekten geschätzt wurde. Allerdings fiel die ermittelte Schnittmenge trotz niedrig-inferenter Operationalisierung geringer aus als erwartet. Demzufolge wurde das Ziel einer möglichst unverzerrten, objektiven Einschätzung der Auftretenshäufigkeit normabweichenden Schülerverhaltens nur bedingt erfüllt.

Dieser Befund ist mit *Grenzen von Umfragen* im Allgemeinen zu erklären. Weil die Beantwortung von Fragebögen auch bei minimiertem Deutungsspielraum immer eine gewisse Interpretationsleistung erfordert, welche zwangsläufig subjektiv geprägt ist. Diese grundsätzliche Limitation wurde bei der Konzeptualisierung der SUGUS-Studie unterschätzt. Rückblickend lassen sich die hohen Rater-Effekte mit wahrnehmungspsychologischen und mit methodologischen Argumenten erklären. Vermutlich hat es die Befragten in einigen Fällen auch überfordert, dass sie die konkrete Anzahl Normabweichungen über den Zeitraum von zwei Wochen einschätzen sollten, sodass ihre Ratings zum Teil auch einem Bauchgefühl entsprechen.

(ii) *Qualität der Stichprobe.* Die Teilnahme an der SUGUS-Studie war komplett freiwillig (mit Ausnahme einer Lehrperson, welche gemäss eigener Aussage von ihrer Schulleitung angewiesen wurde, an der Studie teilzunehmen). Die *Anmeldequote* der angefragten Lehrpersonen beläuft sich wahrscheinlich auf unter 20% – exakt bestimmen lässt sie sich aufgrund der regional unterschiedlich ablaufenden Rekrutierung nicht (Eckstein et al., 2018). Eine Ursache für diese geringe Quote dürfte die gegenwärtige starke Beforschung des Schulfeldes darstellen. Es stellt sich somit die Frage, welche Lehrpersonen mit welchen Eigenschaften sich überhaupt angemeldet hatten. Waren es vielleicht die Übermotivierten? Oder waren es nur Lehrpersonen mit geringen Belastungen?

Darüber hinaus war auch die Teilnahme für die Schüler/innen freiwillig. *275 Kinder haben sich von der Studie abgemeldet*, sodass über sie nur Einschätzungen der Lehrperson vorliegen (keine Selbsteinschätzungen, keine Peer-Ratings). Just diese Kinder verhalten sich nach Ansicht der Lehrpersonen häufiger deviant als diejenigen, welche sich zur Studie angemeldet hatten ($t = 3.56$; $df = 314.17$; $p < .001$). Diese Mittelwertsdifferenz ist aber nur als schwacher Effekt ($d = .31$) zu beurteilen (Cohen, 1988), was die Befürchtung möglicher Verzerrungen relativiert.

Bei den Nicht-Teilnehmenden mussten aufgrund der fehlenden Peer- und Self-Ratings bei der Analyse der manifesten Daten (z.B. bei den klassenspezifischen Auswertungen, ANOVAs) auch die Teacher-Ratings listenweise ausgeschlossen werden. Im SEM Kontext wurden fehlenden Peer- und Self-Ratings anhand aller verfügbaren Informationen nach einer Full Information Maximum Likelihood (FIML) Prozedur geschätzt (Muthén & Muthén, 2017). Dieses Verfahren ist zulässig, weil die fehlenden Werte nach Enders und Bandalos (2001) als Missing at Random (MAR) aufgefasst werden können.

Die Selbstselektion der Schüler/innen und der Lehrpersonen stellen die *Repräsentativität* der Stichprobe infrage. Allerdings ist Repräsentativität "in der Forschungspraxis eher eine theoretische Zielvorgabe" (Bortz & Döring, 2006, S. 398). Um Repräsentativität vollumfänglich zu gewährleisten, bedarf es echter Zufallsstichproben, die sich in der Bildungsforschung gegenwärtig vermutlich nur mit amtlicher Teilnahmeverpflichtung für Lehrpersonen und Schüler/innen realisieren lassen. Einige Befunde sprechen dafür, ein strukturtreues Sample gezogen zu haben, z.B. hinsichtlich der Häufigkeit, mit der die SUGUS Lehrpersonen bestimmte Lehr-Lernformen inszenieren (Eckstein et al., 2018). Ganz sicher ist es aber nicht, ob die vorliegende Stichprobe die Population in jeder Hinsicht genau widerspiegelt. Somit lässt sich nicht mit letzter Sicherheit klären, inwieweit die anhand der untersuchten Stichprobe ermittelten Befunde, z.B. das durchschnittlich geringe Problemausmass, auch auf die Population zutreffen.

(iii) *Spannungsfeld Theorie – Empirie*. Die in den Beiträgen II und IV ermittelten hohen Korrelationen zwischen den Konstrukten "Devianz" und "Störungsempfinden" in der von allen Ratern intersubjektiv geteilten Sichtweise (L2) stellen die theoretische Entflechtung der Merkmalsdimensionen gestörten Unterrichts ein Stück weit in Frage. Die Resultate legen nahe, dass sich die Konstrukte – unter Ausschluss der subjektiven, raterspezifischen Anteile – nicht klar unterscheiden lassen. In der (kleinen) gemeinsamen Sichtweise von Lehrpersonen, Schülerinnen und Schülern vermengen sich die Konstrukte zu einem *homogenen Störungserlebnis*. Immerhin lässt sich die theoretisch hergeleitete Unterscheidung der Konstrukte unter Berücksichtigung der (substanziellen) Rater-Effekte auch empirisch nachzeichnen. Doch es bleibt festzuhalten, dass das Verhältnis von Theorie zu Empirie in dieser Hinsicht kompliziert, vielschichtig und nicht ganz widerspruchsfrei ist.

(iv) *Bislang unerfüllte Ertragserwartungen.* Als das SUGUS Projekt im Jahr 2013 beim Schweizerischen Nationalfonds beantragt wurde, standen bei den Hauptfragestellungen die Bedingungen von Devianz und Störungsempfinden im Vordergrund, etwa didaktische Konfigurationen des Unterrichts. Die bisherigen Analysen tragen aber noch nichts zur Klärung dieser spezifischen Fragen bei. Ein Grund dafür war ein unvorhergesehener Mehraufwand bei der Nachbereitung des Standard Pretests: Der ursprüngliche Zeitplan sah dafür sechs Monate vor, sodass die Hauptstudie bereits im Jahr 2015 hätte durchgeführt werden können. Doch dann hat sich die Auswertung der Pretest-Daten als anforderungsreicher herausgestellt als gedacht. Darüber hinaus wurden die theoretischen Grundlagen des Projekts nach der Vorstudie im Zuge eines längeren Review-Prozesses von Beitrag II der Dissertation (Eckstein, Grob, et al., 2016) substanziell erweitert und vertieft.

Ein positiver Aspekt dieser Verzögerung besteht darin, dass grundlegende Themen theoretischer und methodischer Art intensiver bearbeitet wurden als geplant. Die mit den bisherigen Analysen geklärten Fragen sind thematisch an der Schnittstelle von Pädagogik, Wahrnehmungspsychologie, Erkenntnistheorie und Forschungsmethoden angesiedelt. Das heisst: Mit der vorliegenden Dissertation wurde ein Stück weit *anwendungsorientierte Grundlagenforschung* betrieben. Auf diesem Fundament lassen sich nun die praxisrelevanten Fragen zu den Ursachen gestörten Unterrichts bearbeiten.

3.4. Ausblick auf weitere Analysen im SUGUS-Projekt

Meine Klasse kommt bei diesem Fragebogen sehr gut weg, was auch daran liegt, dass es bei den SuS oft nur "schwarz oder weiss" gibt. Damit meine ich, dass es ihnen beim Auftrag Stillarbeit oft gut gelingt nicht zu stören. Bei offenen Unterrichtsformen oder Partner- & Gruppenarbeiten überbietet es dann aber rasch. Das kommt nicht so zur Geltung bei der Auswertung. Sie "funktionieren" wirklich sehr gut, wenn "der Polizist" schaut & es ganz still sein muss, sonst sind sie aber extreme "Schwatzis" :), nicht zuletzt, weil sie es miteinander so gut haben... (Klasse 023, Lehrperson)

Wenn Kinder Grimassen machen oder bei Stillarbeit schwatzen und wier deswegen nicht in ruhe arbeiten können finde ich das ziemlich blöd. (Klasse 061, Schüler/in 17)

Ich habe es überhaupt nicht gerne, wenn ein Kind (oder mehrere) in einer Gruppenarbeit sich nicht konzentriert. Darum arbeite ich lieber alleine. (Klasse 049, Schüler/in 08)

Bei diesen Zitaten handelt es sich um die abschliessenden Kommentare einer Lehrperson und zwei Kindern, die sich an der SUGUS Hauptstudie beteiligten. Die Aussagen passen zu den SUGUS Hypothesen, wonach der *didaktische Kontext* sowohl die Auftretenshäufigkeit unterrichtlicher Devianz als auch das subjektive Störungsempfinden beeinflusst. Dass selbst Kinder diesen mutmasslichen Einfluss bewusst wahrnehmen und beschreiben, verdeutlicht die Relevanz weiterführender Analysen in diese Richtung.

Die *ursprüngliche Auswertungsstrategie* der quantitativen SUGUS-Studie sah hierfür vor, das in Beitrag IV (Eckstein, under review) erarbeitete Zwei-Ebenen Modell (Abbildung 14) zu einem *Pfadmodell mit einer dritten Ebene* zu weiterzuentwickeln: Zusätzlich zur Ebene der unigen Peer-Ratings (L1) und zur Zielschüler-Ebene (L2) hätte die Ebene der Lehrperson bzw. der Klasse (L3) explizit modelliert werden sollen. Die didaktische Unterrichtspraxis sowie Personenmerkmale der Lehrpersonen hätten auf L3 modelliert werden sollen. Abbildung 15 illustriert das geplante Modell skizzenhaft. Die Abbildung stellt keinen elaborierten Plan dar, es handelt sich lediglich um eine grobe Veranschaulichung der ursprünglich geplanten Analyse.

Level 3: classes / teachers**Guter Unterricht (Tiefenstruktur)**

- Hoher Grad an Binnendifferenzierung
- Klarheit und Strukturiertheit
- Warmes Klassenklima

Inszenierung (Oberflächenstrukt.)

- Häufiger Frontalunterricht

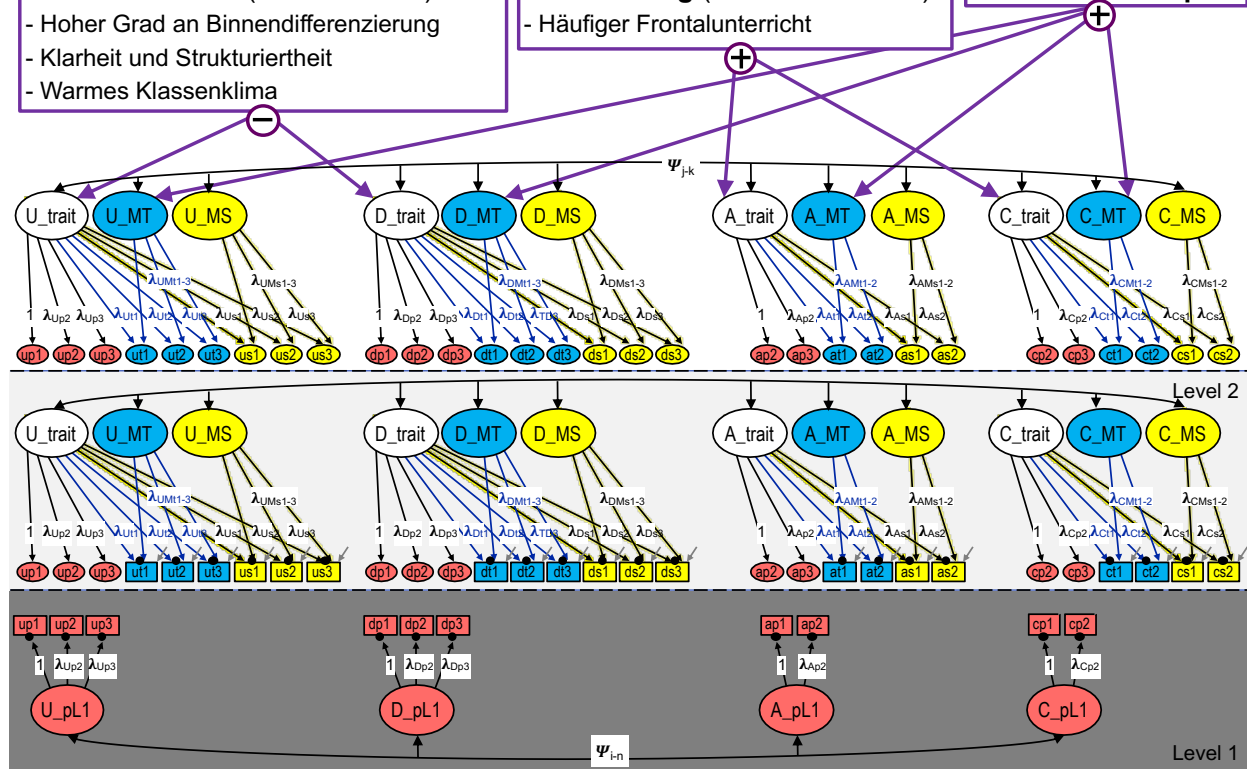
Sensitivität Lp

Abbildung 15. Skizze des ursprünglich geplanten Drei-Ebenen-Modells.

Anmerkungen: Eine detaillierte Erläuterung der Modellspezifikation bezüglich L1 und L2 findet sich in den Anmerkungen zu Abbildung 14. Neu sollten die zielschülerspezifischen Angaben auch auf der Klassenebene (L3) aggregiert werden. Auf L3 sind auch mutmassliche Prädiktoren angesiedelt (violett umrahmte Kästen). Die violetten Pfeile illustrieren holzschnittartig die vermuteten Effekte. Das Vorzeichen am oberen Pfeilkopf illustriert die angenommene Richtung der Effekte (z.B. häufige Binnendifferenzierung verringert die Auftretenshäufigkeit unterrichtlicher Devianz).

Allerdings hat diese ursprüngliche Strategie *unlösbare Schätzprobleme* verursacht.

Ansichts der Komplexität des vorgesehenen Modells ist vermutlich die Fallzahl von 85

Klassen auf L3 zu klein und/oder die Varianz der auf Klassenebene aggregierten Zielschüler-

Daten ist zu gering. Zudem besteht ein grundsätzliches Problem darin, dass *gerichtete Effekte*

exogener Variablen nicht simultan auf Trait- und Methoden-Faktoren der selben Trait-Method-

Einheit bezogen werden dürfen, weil diese per Definition unkorreliert sind (Koch, Holtmann,

Bohn & Eid, 2017). Es könnte aber sein, dass beispielsweise ein Unterrichtsmerkmal die intersubjektiv geteilte Einschätzung der Devianz (Trait-Faktor) sowie die individuelle Abweichung davon bei den Teacher-Ratings (Methoden-Faktor) beeinflusst. Diese Annahme nach dem CT-C(M-1) Ansatz korrekt zu testen, wäre äusserst anspruchsvoll und zeitaufwändig.

Aus diesen Gründen werden gegenwärtig folgende *vier Massnahmenpakete* in Betracht gezogen, um die Komplexität des Modells zu reduzieren, sodass die mutmasslichen Einflüsse der Didaktik und der Lehrperson doch noch korrekt analysiert werden können:

(1) *Reduktion der Methoden.* Auf die Self-Ratings wird komplett verzichtet. Die Devianz-Konstrukte werden nur noch anhand der Peer- und Teacher-Ratings geschätzt. Beim subjektiven Störungsempfinden interessiert einzig die Perspektive der Lehrpersonen, sodass hier nur deren Ratings zur Schätzung herangezogen werden.

(2) *Vereinfachung der Mehrebenenstruktur.* Es wird nur noch mit einer Ebene "flach" gerechnet. Dazu werden die multiplen Peer-Ratings pro Zielschüler/in (L1) manifest aggregiert und als messfehlerbehaftete Indikatoren neben den Teacher-Ratings modelliert. Die Verschachtelung mehrerer Zielschüler/innen pro Klasse wird mit der Mplus implementierten Option "COMPLEX" (Muthén & Muthén, 2017) berücksichtigt, um die Standardfehler nicht zu unterschätzen.

(3) *Abkehr vom CT-C(M-1) Ansatz.* Die Devianz-Inzidenz wird weiterhin im Sinne eines Trait-Faktors geschätzt, denn hier interessiert die gemeinsame Sichtweise von Peers und Lehrpersonen. Neu wird aber die raterspezifische Varianz nicht mehr explizit in Methoden-Faktoren modelliert, sondern schlicht als Teil der Residualvarianz in den Messfehlern der

Indikatoren belassen (wie in vielen Analysen üblich). Um die Spezifität der Teacher-Ratings bzw. der Peer-Ratings zu berücksichtigen, dürfen die Residuen pro Methode miteinander korrelieren. Diese Korrelationen ersetzen die fehlenden Methoden-Faktoren.

(4) *Zusammenfassung der Subdimensionen des Konstrukts Störungsempfinden.* Die beiden bislang separat modellierten Konstrukte affektives und kognitives Störungsempfinden werden zu einem Superkonstrukt "allgemeines Störungsempfinden" zusammengefasst.

Vorläufige Analysen legen nahe, dass diese Strategie aufgehen dürfte. Abbildung 16 illustriert ein Pfadmodell, welches gut zu den Daten passt ($\chi^2[\text{MLR}] = 256.48$, $df = 152$, $p < .001$; RMSEA = .02; CFI = .97; SRMR = .04). Die *Ergebnisse* dieses vorläufigen Modells zeigen erwartungsgemäss, dass das zielschülerspezifische Störungsempfinden der Lehrpersonen (PD_T) mit $\beta = .78^*$ hochgradig davon abhängig ist, wie häufig sich die betreffenden Zielschüler/innen undiszipliniert verhalten – gemessen an der intersubjektiv geteilten Einschätzung (U_trait). Zudem konnte erstmals ein Effekt ($\beta = .19^*$) der *generellen Sensitivität* der Lehrpersonen (GS_trait) auf ihr zielschülerspezifisches Störungsempfinden nachgewiesen werden. Dieser Effekt bedeutet: Inwieweit die Lehrpersonen ihre Schüler/innen als störend empfinden, hängt auch zu einem (geringfügigen) Teil von ihrer allgemeinen Empfindlichkeit ab – unabhängig davon, wie häufig sich die Schüler/innen im Unterricht undiszipliniert verhalten. Darüber hinaus wurde auch erstmals ein (kleiner) Effekt ($\beta = -.09^*$) des *Kooperationsgrads* (COOP) auf das Störungsempfinden der Lehrperson nachgewiesen: In unterrichtlichen Settings, in denen häufig Gruppen- und Partnerarbeit sowie selten Einzelarbeit praktiziert wird, fällt das Störungsempfinden leicht tiefer aus als in weniger kooperativen Settings. Diese Variable hat jedoch keinen Einfluss auf die Auftretenshäufigkeit undisziplinierten Verhaltens (U_trait).

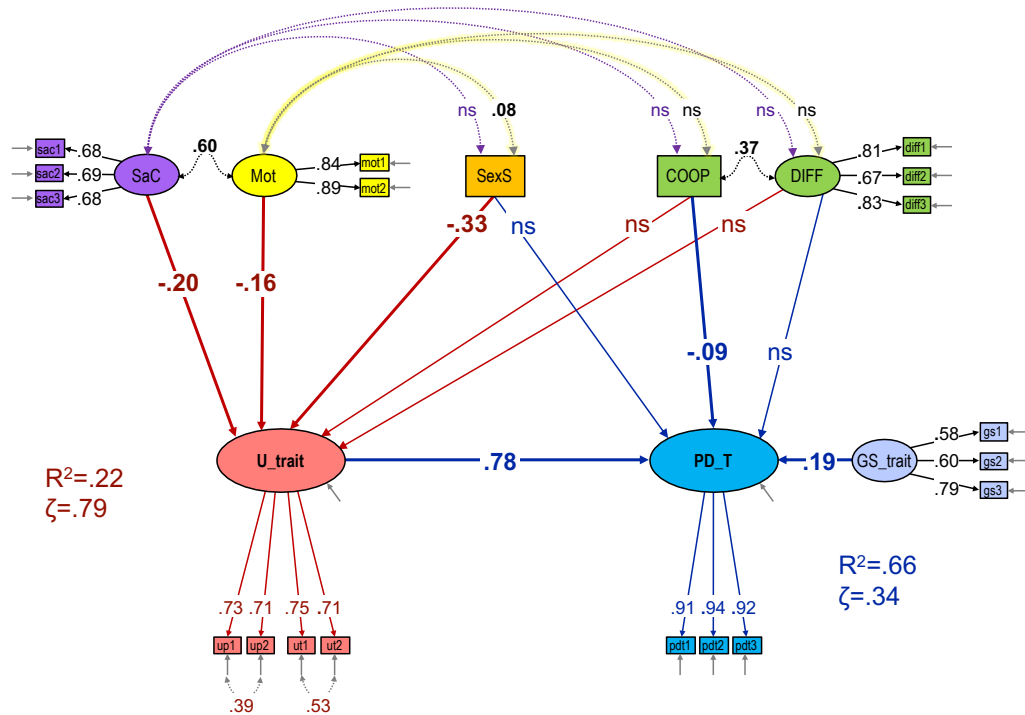


Abbildung 16. Ergebnisse des Pfadmodells nach neuer Analysestrategie.

Anmerkungen: Abgebildet sind die standardisierten Parameter, die alle signifikant sind mit $p < .05$, sodass auf die übliche Kennzeichnung mit Asterisk verzichtet wurde. Nicht eingezeichnete Korrelationen (z.B. zwischen SexS und DIFF) sind aus theoretischen Gründen auf den Wert 0 restringiert worden.

U_trait = Trait-Faktor Undiszipliniertes Verhalten (zielschülerspezifisch), geschätzt aus den manifest aggregierten Peer-Ratings (up1, up2) sowie aus den Teacher-Ratings (ut1–ut2)

PD_T = Zielschülerspezifisches Störungsempfinden der Lehrperson (affektiv und kognitiv)

GS_trait = Generelle Störungssensitivität der Lehrpersonen

SaC = Strukturiertheit und Klarheit, geschätzt aus den Angaben der in U_trait und PD_T beurteilten Zielschülerinnen und Zielschülern; sac1-sac3 sind Item-Parcels aus total 6 Items (z.B. "Ich kann dem Unterricht gut folgen")

Mot = Motivation, geschätzt aus den Zielschülerangaben; mot1 und mot2 sind Item-Parcels aus total 4 Items (z.B. "Ich finde den Unterricht sehr spannend")

SexS = Geschlecht der/des beurteilten Zielschülerin/Zielschülers (0 = Junge; 1 = Mädchen)

COOP = Kooperationsgrad, geschätzt aus Teacher-Ratings (hohe Werte bedeuten häufige Gruppen- und Partnerarbeit sowie seltene Einzelarbeit)

DIFF = Häufigkeit des Einsatzes binnendifferenzierender Massnahmen, geschätzt aus Teacher-Ratings (hohe Werte bedeuten häufige Differenzierung)

In Bezug auf den Einsatz *binnendifferenzierender Massnahmen* (DIFF) konnte kein Effekt ermittelt werden – weder auf die Auftretenshäufigkeit undisziplinierten Verhaltens (U_trait) noch auf das Störungsempfinden der Lehrpersonen (PD_T). Darüber hinaus ist diese Variable weder mit der Strukturiertheit und Klarheit (SaC) aus Sicht der Zielschüler/innen korreliert, noch mit deren Motivation (Mot). Diese Nulleffekte stehen im Widerspruch zu den theoretischen Annahmen (Kap. 2.2.3), wonach der Differenzierung in allen diesen Beziehungen positive Auswirkungen unterstellt wird. Vermutlich hat dies damit zu tun, dass der Einsatz binnendifferenzierender Massnahmen aus Sicht der Lehrpersonen im Rückblick auf die vergangenen Monate eher grob erfasst wurde. Eine exakte Beschreibung aller Skalen findet sich im technischen Bericht (Eckstein et al., 2018). Eine signifikante Korrelation ($r = .37^*$) wurde zwischen Differenzierung (DIFF) und dem Kooperationsgrad (COOP) ermittelt. Sie besagt, dass in kooperativen Settings vergleichsweise häufig differenziert wird.

Ebenfalls korreliert ($r = .60^*$) sind die Variablen *Strukturiertheit und Klarheit* (SaC) und *Motivation* (Mot). Darüber hinaus haben sie beide einen negativen Einfluss ($\beta = -.20^*$; $\beta = -.16^*$) auf die Auftretenshäufigkeit undisziplinierten Verhaltens (U_trait). Diese Effekte deuten an, dass hohe Unterrichtsqualität störungspräventiv wirkt. Allerdings ist einschränkend zu bemerken, dass die Richtung des Effekts aufgrund der Querschnittsdaten auch umgekehrt sein könnte. Dies würde bedeuten, dass Schüler/innen dem Unterricht besser folgen können und sich durch eine höhere Lernmotivation auszeichnen, wenn sie sich selten undiszipliniert verhalten.

Der negative Einfluss ($\beta = -.33^*$) des Geschlechts der Zielschüler/innen (SexS) auf die Variable U_trait bedeutet, dass sich Mädchen seltener undiszipliniert verhalten als Jungen. Die (schwache) Korrelation ($r = .08^*$) mit der Variable SaC bedeutet, dass Mädchen gegenüber Jungen dem Unterricht leicht besser folgen können.

Insgesamt klären die signifikanten Prädiktoren SaC, Mot und SexS gemeinsam rund 22% der Varianz der Auftretenshäufigkeit undisziplinierten Verhaltens (R^2). Rund 79% der Varianz bleiben dagegen unerklärt (ζ). Die Varianz des Kriteriums Störungsempfinden wird von den Prädiktoren zu 66% erklärt (R^2), wobei der Löwenanteil auf das Verhalten der Zielschüler/innen zurückzuführen ist. 34% Varianz bleiben unerklärt (ζ).

Konservativ formuliert zeigt dieses vorläufige Pfadmodell einen Zusammenhang zwischen der Unterrichtsqualität (SaC, Mot) und dem Verhalten der Schüler/innen (U_trait). Aufgrund des Querschnittcharakters der Daten kann dieser Zusammenhang nicht als kausaler Effekt interpretiert werden. Mit theoretischen Argumenten kann jedoch geschlussfolgert werden, dass sich hier *störungspräventive Effekte guten Unterrichts* abzeichnen, die mit weiterführender Forschung untersucht werden sollten. Ergebnisse solcher zukünftigen Studien dürften eine hohe Relevanz haben für die Praxis.

4. Literaturverzeichnis

- Ahrbeck, B. & Willmann, M. (Hrsg.). (2010). *Pädagogik bei Verhaltensstörungen. Ein Handbuch*. Stuttgart: Kohlhammer.
- Allemann-Ghionda, C. (2008). Für die Welt Diversität feiern - im heimischen Garten Ungleichheit kultivieren? Von gegenläufigen Entwicklungen in der Politik, Theorie und Praxis der interkulturellen Bildung in Europa. *Zeitschrift für Pädagogik*, 54(1), 18-33.
- American Psychological Association (Hrsg.). (2010). *Publication manual* (6th ed.). Washington, DC: American Psychological Association.
- Annen, L., Cattaneo, M. A., Denzler, S., Diem, A., Grossenbacher, S., Hof, S. et al. (2010). *Bildungsbericht Schweiz 2010*. Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.
- Arbuckle, C. & Little, E. (2004). Teachers' perceptions and management of disruptive classroom behaviour during the middle years. *Australian Journal of Educational and Developmental Psychology*, 4, 59-70.
- Bakken, J. P., Obiakor, F. E. & Rotatori, A. F. (Hrsg.). (2012). *Behavioral disorders. Practice concerns and students with EBD*. Bingley: Emerald.
- Beaman, R., Wheldall, K. & Kemp, C. (2007). Recent research on troublesome classroom behaviour: A review. *Australasian Journal of Special Education*, 31(1), 45-60.
- Becker, H. S. (1963/1981). *Aussenseiter zur Soziologie abweichenden Verhaltens*. Frankfurt am Main: Fischer Taschenbuch.
- Berg, D., Imhof, M., Kollera, S., Schmidt, U. & Ulber, D. (1998). Häufigkeiten von Verhaltensauffälligkeiten in der Grundschule aus der Sicht der Klassenlehrer. *Psychologie in Erziehung und Unterricht*, 45, 280-290.
- Bernhardt, R. (2007). Störungen, Widerstand und Verweigerung im Unterricht. In D. Katzenbach (Hrsg.), *Vielfalt braucht Struktur Heterogenität als Herausforderung für die Unterrichts- und Schulentwicklung* (S. 149-164). Frankfurt a.M.: Johann Wolfgang Goethe-Universität.
- Bildungsdirektion Kanton Zürich. (2007). *Angebote für Schülerinnen und Schüler mit besonderen pädagogischen Bedürfnissen. Integrative Förderung (IF). Umsetzung Volksschulgesetz*. Verfügbar unter: http://www.vsa.zh.ch/internet/bildungsdirektion/vsa/de/schulbetrieb_und_unterricht/sonderpaedagogisches0/publikationen.html [14.10.2011]
- Bildungsdirektion Kanton Zürich. (2011). Handreichung. Integrative und individualisierende Lernförderung. In B. K. Zürich (Hrsg.).
- Bildungsdirektion Kanton Zürich. (2013). Indikationsbereiche zur Klärung der Indikationen für sonderschulische Massnahmen im Kontext des Standardisierten Abklärungsverfahrens (SAV). Bildungsdirektion des Kantons Zürich.
- Bildungsdirektion Kanton Zürich. (2018). *Klassenführung*. Verfügbar unter: http://www.fsb.zh.ch/internet/bildungsdirektion/fsb/de/schulbeurteilung/rechtliche_grundlagen/qualitaetsbereiche.html [09.03.2018]
- Binswanger, M. (2013a, 15. Juni). Der Bub tut nicht gut: Die Grenzen der Integration. *Tages Anzeiger*.
- Binswanger, M. (2013b, 20. Dezember). Integrieren, bis die Polizei kommt. *Tages Anzeiger*.
- Binswanger, M. (2013c, 17. Juni). Nicht mehr zumutbar. *Tages Anzeiger*.
- Bleich, J. (2013). Wahnsinn wird normal. *Der Spiegel*(4), 111-119.

- Bless, G., Sermier Dessemonet, R. & Benoit, V. (2010). *Evaluation zu den Wirkungen der Integration von Kindern mit einer geistigen Behinderung im Kanton Zürich. Ein Vergleich zwischen Sonderbeschulung, Einzelintegration und Integrationsklassen. Schlussbericht.*: Universität Fribourg.
- Boban, I. & Hinz, A. (2009). Integration und Inklusion als Leitbegriffe der schulischen Sonderpädagogik. In G. Opp, G. Theunissen & J. Teichmann (Hrsg.), *Handbuch schulische Sonderpädagogik* (S. 29-36). Bad Heilbrunn: Klinkhardt.
- Boban, I. & Hinz, A. (2012). Index für Inklusion - auf dem Weg zu einer Schule für alle. In A. Lanfranchi & J. Steppacher (Hrsg.), *Schulische Integration gelingt. Gute Praxis wahrnehmen, Neues entwickeln* (S. 75-88). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Bohl, T. & Kucharz, D. (2010). *Offener Unterricht heute. Konzeptionelle und didaktische Weiterentwicklung.* Weinheim u.a.: Beltz.
- Borders, C., Bock, S. J. & Michalak, N. (2012). Differentiated instruction for students with emotional and behavioral disorders. In J. P. Bakken, F. E. Obiakor & A. F. Rotatori (Hrsg.), *Behavioral disorders. Practice concerns and students with EBD* (S. 203–219). Bingley: Emerald.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4., überarbeitete Auflage). Berlin, Heidelberg: Springer Medizin Verlag Heidelberg.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (Bd. Sonderband 2, S. 183-198). Göttingen: Schwartz.
- Breidenstein, G. (2006). *Teilnahme am Unterricht. Studien zum Schülerjob.* Wiesbaden: Verlag für Sozialwissenschaften.
- Brügelmann, H. (1997). Die Öffnung des Unterrichts muss radikaler gedacht, aber auch klarer strukturiert werden. In H. Balhorn & H. Niemann (Hrsg.), *Sprachen werden Schrift. Mündlichkeit - Schriftlichkeit - Mehrsprachigkeit* (S. 43-57). Lengwil am Bodensee: LibelleVerlag.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler.* Hallbergmoos: Pearson.
- Burri, A. (2017, 21. Mai). Störenfriede raus! *NZZ am Sonntag*.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus. Basic concepts, applications, and programming.* New York, NY: Routledge.
- Carretero-Dios, H., Eid, M. & Ruch, W. (2011). Analyzing multitrait-multimethod data with multilevel confirmatory factor analysis: An application to the validation of the State-Trait Cheerfulness Inventory. *Journal of Research in Personality*, 45(2), 153–164.
- Carroll, A., Houghton, S., Taylor, M., West, J. & List-Kerz, M. (2006). Responses to interpersonal and physically provoking situations. *Educational Psychology*, 26(4), 483–498.
- Casale, G., Grosche, M., Volpe, R. J. & Hennemann, T. (2017). Zuverlässigkeit von Verhaltensverlaufsdiagnostik über Rater und Messzeitpunkte bei Schülern mit externalisierenden Verhaltensproblemen. *Empirische Sonderpädagogik*, 9(2), 143–164.
- Christe, G. (2010). Mechanismen der Ausgrenzung von Jugendlichen im Bildungssystem. In H. Ricking (Hrsg.), *Förderbedarf in der emotionalen und sozialen Entwicklung Prävention, Interdisziplinarität und Professionalisierung* (S. 41-52). Bad Heilbrunn: Klinkhardt.

- Cialdini, R. B. & Reno, R. R. (1990). A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. *Journal of Personality and Social Psychology*, 58(6), 1015-1026.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd). Hillsdale, NJ: Erlbaum.
- Coxe, S., West, S. G. & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of personality assessment*, 91(2), 121-136.
- Crawshaw, M. (2015). Secondary school teachers' perceptions of student misbehaviour: A review of international research, 1983 to 2013. *Australian Journal of Education*, 59(3), 293-311.
- Croci, A., Imgrüth, P., Landwehr, N. & Spring, K. (1995). *ELF: Ein Projekt macht Schule : Magazin zum Thema erweiterte Lernformen* (1). Littau, Buchs: Kantonaler Lehrmittelverlag Luzern Kantonaler Lehrmittelverlag Aargau.
- de Boer, A., Pijl, S. J. & Minnaert, A. (2011). Regular Primary Schoolteachers' Attitudes Towards Inclusive Education: A Review of the Literature. *International Journal of Inclusive Education* 15(3), 331-353.
- Dellwing, M. (2011). Truth in labeling: Are descriptions all we have? *Deviant Behavior*, 32(7), 653-675.
- Deluigi, T. (2016). Das ideale und das abweichende Schulkind: Pädagogisches Wissen für die Ausbildung einer Profession. In M. Götz & M. Vogt (Hrsg.), *Schulwissen für und über Kinder. Beiträge zur historischen Primarschulforschung* (S. 58-75). Bad Heilbrunn: Klinkhardt.
- Deutsche Gesellschaft für Psychologie (Hrsg.). (2007). *Richtlinien zur Manuskriptgestaltung* (3. ed.). Göttingen: Hogrefe.
- Dewey, J. (1916/1994). *Erziehung durch und für Erfahrung. 2. Auflage*. Stuttgart: Klett-Cotta.
- Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A. & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, 106(2), 569-583.
- Dlugosch, A. (2010). Selbst, Identität, Entwicklung und Krisen. In B. Ahrbeck & M. Willmann (Hrsg.), *Pädagogik bei Verhaltensstörungen. Ein Handbuch* (S. 113-120). Stuttgart: Verlag W. Kohlhammer.
- Donzé, R. (2016, 20. März). Schiffbruch einer schönen Idee. *NZZ am Sonntag*.
- Dörpinghaus, A. (2009). Pathologisierungen schulischen Verhaltens. Über die biopolitische Umschrift pädagogischer Deutungsmuster. Ein Essay. In A. Nießeler & I. K. Uphoff (Hrsg.), *Pädagogische Auffälligkeiten. Deutungsmuster von Verhaltensstörungen und Verhaltensauffälligkeiten - kritisch betrachtet* (S. 15-30). Würzburg: Königshausen & Neumann.
- Doyle, W. (2006). Ecological approaches to classroom management. In C. Evertson & C. Weinstein (Hrsg.), *Handbook of classroom management: Research, practice and contemporary issues* (S. 97-125). Mahwah, NJ: Erlbaum.
- Eckstein, B. (2018). Unterrichtsstörungen: eine Frage der Perspektive? In S. Schwab, G. Tafner, S. Luttenberger, H. Knauder & M. Reisinger (Hrsg.), *Von der Wissenschaft in die Praxis? Zum Verhältnis von Forschung und Praxis in der Bildungsforschung* (S. 78-92). Münster: Waxmann.

- Eckstein, B. (under review). Production and Perception of Classroom Disturbances – A new approach to investigating the perspectives of teachers and students.
- Eckstein, B., Grob, U. & Reusser, K. (2015). Unterrichtliche Devianz und subjektives Störungsempfinden. Entwicklung eines theoretisch fundierten Instrumentariums zur Erfassung von Unterrichtsstörungen, *AESF-Frühjahrstagung*. Potsdam (DE).
- Eckstein, B., Grob, U. & Reusser, K. (2016). Unterrichtliche Devianz und subjektives Störungsempfinden. Entwicklung eines Instrumentariums zur Erfassung von Unterrichtsstörungen. *Empirische Pädagogik*, 30(1), 113–129.
- Eckstein, B., Grob, U. & Reusser, K. (2017). Production and Perception of Classroom Disturbances. Personal and Contextual Facets, *EARLI Biennial Main Conference. Education in the Crossroads of Economy and Politics*. Tampere (FIN).
- Eckstein, B., Luger, S., Grob, U. & Reusser, K. (2016a). SUGUS – Studie zur Untersuchung gestörten Unterrichts. Ergebnisbericht der Hauptstudie – anonymisierte Fassung. Universität Zürich. Verfügbar über www.ife.uzh.ch/SUGUS [14.2.2017].
- Eckstein, B., Luger, S., Grob, U. & Reusser, K. (2016b). SUGUS – Studie zur Untersuchung gestörten Unterrichts. Ergebnisbericht der Hauptstudie in 85 unveröffentlichten, personalisierten Fassungen zuhanden der teilnehmenden Lehrpersonen. Universität Zürich.
- Eckstein, B., Luger, S., Grob, U. & Reusser, K. (2018). *SUGUS: Technischer Bericht der quantitativen Teilstudie. Studiendesign, Stichprobe und Skalendokumentation*.
- Eckstein, B., Reusser, K., Grob, U. & Hofstetter, A. (2015a). *SUGUS – Studie zur Untersuchung gestörten Unterrichts. Kurzer Ergebnisbericht der Vorstudie – anonymisierte Fassung*. Zürich: Universität Zürich.
- Eckstein, B., Reusser, K., Grob, U. & Hofstetter, A. (2015b). SUGUS – Studie zur Untersuchung gestörten Unterrichts. Kurzer Ergebnisbericht der Vorstudie in elf unveröffentlichten, personalisierten Fassungen zuhanden der teilnehmenden Lehrpersonen. Universität Zürich.
- Eckstein, B., Reusser, K., Stebler, R. & Mandel, D. (2013). Umsetzung der integrativen Volksschule – Was Lehrpersonen optimistisch macht: Eine Analyse der Überzeugungen von Klassenlehrpersonen im Kanton Zürich. *Schweizerische Zeitschrift für Bildungswissenschaften*, 35(1), 91–112.
- Eder, F. (2006). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie*. 3., überarb. und erw. Aufl (S. 622–631). Weinheim u.a.: Beltz PVU.
- Eid, M., Lischetzke, T., Nussbeck, F. & Trierweiler, L. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C (M-1) model. *Psychological Methods*, 8(1), 38–60.
- Eid, M., Nussbeck, F., Geiser, C., Cole, D. A., Gollwitzer, M. & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, 13(3), 230–253.
- Emmer, E. T. & Evertson, C. M. (2013). *Classroom management for middle and high school teachers*. Boston; Munich: Pearson Education.
- Emmer, E. T. & Sabornie, E. J. (2015). *Handbook of classroom management* (2nd). New York, NY: Routledge.
- Enders, C. K. & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling*, 8(3), 430–457.

- Finney, S. J. & DiStefano, C. (2013). Nonnormal and categorical data in structural equation models. In G. R. Hancock & R. O. Mueller (Hrsg.), *Structural equation modeling: A second course* (2nd ed., S. 439–492). Charlotte, NC: Information Age Publishing.
- Fishbein, M. & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.
- Fröhlich-Gildhoff, K. (2007). *Verhaltensauffälligkeiten bei Kindern und Jugendlichen*. Stuttgart: Kohlhammer.
- Frölich, J., Döpfner, M. & Banaschewski, T. (2014). *ADHS in Schule und Unterricht. Pädagogisch-didaktische Ansätze im Rahmen des multimodalen Behandlungskonzepts*. Stuttgart: Kohlhammer.
- Gebhardt, M., Schwab, S., Reicher, H., Ellmeier, B., Gmeiner, S., Rossmann, P. et al. (2011). Einstellungen von LehrerInnen zur schulischen Integration von Kindern mit einem sonderpädagogischen Förderbedarf in Österreich. *Empirische Sonderpädagogik*(4), 275–290.
- Geiser, C., Eid, M., Nussbeck, F., Lischetzke, T. & Cole, D. A. (2010). Multitrait-Multimethod-Analyse. In H. Holling & B. Schmitz (Hrsg.), *Handbuch Statistik, Methoden und Evaluation* (S. 679–685). Göttingen: Hogrefe.
- Gentz, D. (2012). Verhaltensmodifikation von Schülern - das Smiley-System. In A. Ohidy (Hrsg.), *Heterogenität und Lehrerhandeln im Spiegel erziehungswissenschaftlicher Fallstudien* (S. 315–333). Baltmannsweiler: Schneider Verlag Hohengehren.
- Gest, S. D. & Rodkin, P. C. (2011). Teaching practices and elementary classroom peer ecologies. *Journal of Applied Developmental Psychology*, 32(5), 288–296.
- Glötzl, H. (1976). *Die Produktion abweichenden Verhaltens in der Volksschule. Labeling Approach und abweichendes Schülerverhalten*. Regensburg.
- Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S. et al. (2016). Off-task behavior in elementary school children. *Learning and Instruction*, 44, 128–143.
- Göppel, R. (2010). Von der "sittlichen Verwilderung" zu "Verhaltensstörungen" - Zur Begriffs- und Ideengeschichte der pädagogischen Reflexion über "schwierige Kinder". In B. Ahrbeck & M. Willmann (Hrsg.), *Pädagogik bei Verhaltensstörungen. Ein Handbuch* (S. 11–20). Stuttgart: Kohlhammer.
- Göppel, R. (2014). *Gehirn, Psyche, Bildung. Chancen und Grenzen einer Neuropädagogik*. Stuttgart: Verlag W. Kohlhammer.
- Granot, D. (2014). Teacher-student attachment and student school adaptation: A variable centered and person centered analytical approaches. *American Journal of Educational Research*, 2(11), 1005–1014.
- Grob, U. & Maag Merki, K. (2001). *Überfachliche Kompetenzen theoretische Grundlegung und empirische Erprobung eines Indikatorensystems*. Dissertation, Lang, Bern.
- Grosse Siestrup, C. (2008). *Unterrichtsstörungen aus der Sicht von Lehrenden und Lernenden*. Frankfurt: Lang.
- Grünke, M. (2008). Offener Unterricht und Projektunterricht. In M. Fingerle & S. Ellinger (Hrsg.), *Sonderpädagogische Förderprogramme im Vergleich* (S. 13–33). Stuttgart: Verlag W. Kohlhammer.
- Haag, L. & Streber, D. (2012). *Klassenführung. Erfolgreich unterrichten mit Classroom Management*. Weinheim: Beltz.

- Hamre, B. K., Pianta, R. C., Downer, J. T. & Mashburn, A. J. (2008). Teachers' perceptions of conflict with young students: Looking beyond problem behaviors. *Social Development*, 17(1), 115–136.
- Hargreaves, D. H., Hester, S. K. & Mellor, F. J. (1981). *Abweichendes Verhalten im Unterricht*. Weinheim; Basel: Beltz.
- Harrison, J., Vannest, K., Davis, J. & Reynolds, C. (2012). Common problem behaviors of children and adolescents in general education classrooms in the United States. *Journal of Emotional and Behavioral Disorders*, 20(1), 55–64.
- Hartke, B., Diehl, K., Mahlau, K. & Voss, S. (2014). Prävention und Integration im Anschluss an den Response-to-Intervention-Ansatz (RTI): Das Rügener Inklusionsmodell (RIM). In K. Popp (Hrsg.), *Schülerinnen und Schüler mit herausforderndem Verhalten. Hilfen für die schulische Praxis* (S. 89-100). Stuttgart: Kohlhammer.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts. Franz Emanuel Weinert gewidmet. Neubearb., 1. Aufl.* Seelze-Velber: Kallmeyer u.a.
- Hempel-Jorgensen, A. (2009). The construction of the “ideal pupil” and pupils' perceptions of “misbehaviour” and discipline: Contrasting experiences from a low-socio-economic and a high-socio-economic primary school. *British Journal of Sociology of Education*, 30(4), 435–448.
- Hennemann, T., Hövel, D., Casale, G., Hagen, T. & Fitting-Dahlmann, K. (2015). *Schulische Prävention im Bereich Verhalten*. Stuttgart: Kohlhammer.
- Hightower, A., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. et al. (1986). The Teacher–Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review*, 15(3), 393-409.
- Hilbe, J. M. (2011). *Negative binomial regression* (2nd). Cambridge: Cambridge University Press.
- Hilbe, J. M. (2014). *Modeling count data*. New York, NY: Cambridge University Press.
- Hillenbrand, C. (2011). *Didaktik bei Unterrichts- und Verhaltensstörungen* (3., aktualisierte Aufl.). München: Reinhardt.
- Hofer, M. (1986). *Sozialpsychologie erzieherischen Handelns [Social psychology of education]*. Göttingen: Hogrefe.
- Hofstetter, A. & Reusser, K. (2016). Unterrichtsstörungen und deren Prävention – Konzepte eines heterogenitätsorientierten Unterrichts als Chancen oder Herausforderungen? Unveröffentlichte Literaturarbeit. Zürich: Universität Zürich. Institut für Erziehungswissenschaft.
- Hofstetter, A. & Reusser, K. (2018). Wer oder was stört hier eigentlich? Eine Untersuchung des Störungsempfindens von Lernenden und Lehrenden während unterschiedlicher unterrichtlicher Aktivitätsstrukturen, *SGBF Kongress. Bildung – Politik – Staat*. Zürich.
- Höhn, E. (1967/1980). *Der schlechte Schüler. Sozialpsychologische Untersuchungen über das Bild des Schulversagers. Überarbeitete Neuauflage*. München: Piper.
- Holtappels, H. G. (2000). "Abweichendes Verhalten" und soziale Etikettierungsprozesse in der Schule. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Pädagogisch-psychologische Aspekte des Lehrens und des Lernens in der Schule* (S. 231-256). Opladen: Leske u. Budrich.

- Homfeldt, H. G. (1974). *Stigma und Schule. Abweichendes Verhalten bei Lehrern und Schülern*. Düsseldorf: Pädagogischer Verlag Schwann.
- Hox, J. (2010). *Multilevel analysis: techniques and applications* (2nd). New York, NY: Routledge.
- Hoyt, W. T. & Kerns, M.-D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, 4(4), 403–424.
- Hunt, D., Carline, J., Tonesk, X., Yergan, J., Siever, M. & Loebel, J. (1989). Types of problem students encountered by clinical teachers on clerkships. *Medical Education*, 23(1), 14–18.
- Hurrelmann, K., Klocke, A., Melzer, W. & Ravens-Sieberer, U. (2003). *WHO-Jugendgesundheitsurvey. Konzept und ausgewählte Ergebnisse für die Bundesrepublik Deutschland*. Weinheim: Juventa-Verlag.
- Ihle, W. & Esser, G. (2002). Epidemiologie psychischer Störungen im Kindes- und Jugendalter: Prävalenz, Verlauf, Komorbidität und Geschlechtsunterschiede. *Psychologische Rundschau*, 53(4), 159–169.
- Infantino, J. & Little, E. (2005). Students' perceptions of classroom behaviour problems and the effectiveness of different disciplinary methods. *Educational Psychology*, 25(5), 491–508.
- Jak, S., Oort, F. J. & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling*, 20(2), 265–282.
- Jang, H., Reeve, J. & Deci, E. L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology*, 102(3), 588–600.
- Jaumann-Graumann, O. (2000). Offener Unterricht - ja, aber strukturiert. *Grundschule*, 9, 36–38.
- Jenzer, C. (1998). *Schulstrukturen als historisch gewachsenes Produkt bildungspolitischer Vorstellungen. Blitzlichter in die Entstehung der schweizerischen Schulstrukturen*. Bern u.a.: Lang.
- Joller-Graf, K. (2010). Binnendifferenziert unterrichten. In A. Buholzer & A. Kummer-Wyss (Hrsg.), *Alle gleich - alle unterschiedlich! Zum Umgang mit Heterogenität in Schule und Unterricht*. Seelze-Velber: Kallmeyer.
- Kantonsrat des Kantons Zürich. (2005). Volksschulgesetz (VSG).
- Katzenbach, D. (2017). Inklusion und Heterogenität. In *Umgang mit Heterogenität in Schule und Unterricht: Grundlagentheoretische Beiträge und didaktische Reflexionen* (S. 123–139).
- Keller, G. (2014). *Die Schülerschelte. Leidensgeschichte einer Generation*. Herbolzheim: Centaurus Verlag & Media UG.
- Kleinke, K., Schlüter, E. & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus. Eine praktische Einführung [Structural equation modelling with Mplus. A practical introduction]* (2nd). Berlin: de Gruyter.
- Klieme, E. & Warwas, J. (2011). Konzepte der individuellen Förderung. *Zeitschrift für Pädagogik*, 57(6), 805–818.
- Koch, T., Holtmann, J., Bohn, J. & Eid, M. (2017). Explaining General and Specific Factors in Longitudinal, Multimethod, and Bifactor Models: Some Caveats and Recommendations. *Psychological Methods*.
- Kokkinos, C. M. (2007). Job stressors, personality and burnout in primary school teachers. *British Journal of Educational Psychology*, 77, 229–243.

- Korsch, F. & Petermann, F. (2012). Früherkennung von Verhaltensstörungen durch die schulärztliche Eingangsuntersuchung. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 61(9), 691–705.
- Kounin, J. S. (1976/2006). *Techniken der Klassenführung* ([Neuaufg.]). Münster: Waxmann.
- Krause, A., Dorsemagen, C. & Alexander, T. (2011). Belastung und Beanspruchung im Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 788-813). Münster: Waxmann.
- Kreuzberger, N. (2002). *Aufmerksamkeitsverhalten in offener Lernsituation in der Sekundarstufe. Untersuchungen zur Öffnung von Unterricht am Gymnasium*. Zugl Bielefeld, Univ , Diss , 2002, Hänsel-Hohenhausen, Frankfurt am Main.
- Krumm, V. (2003). *Wie Lehrer ihre Schüler disziplinieren. Ein Beitrag zur „Schwarzen Pädagogik“*. Verfügbar unter: <http://www.lernwelt.at/downloads/wielehrerihreschuelerdisziplinieren.pdf> [10.10.2012]
- Krumm, V. & Weiß, S. (2006). Ungerechte Lehrer. Zu einem Defizit in der Forschung über Gewalt an Schulen. In W. Melzer (Hrsg.), *Gewalt an Schulen. Analyse und Prävention*. (S. 123-146).
- Kuester, D. A. & Zentall, S. S. (2012). Social interaction rules in cooperative learning groups for students at risk for ADHD. *The Journal of Experimental Education*, 80(1), 69–95.
- Kunz, A., Luder, R. & Moretti, M. (2010). Die Messung von Einstellungen zur Integration (EZI). *Empirische Sonderpädagogik*(3), 83-94.
- Kyriacou, C., Avramidis, E., Hoie, H., Stephens, P. & Hultgren, A. (2007). The development of student teachers' views on pupil misbehaviour during an initial teacher training programme in England and Norway. *Journal of Education for Teaching: International Research and Pedagogy*, 33(3), 293–307.
- Lamnek, S. (2007). *Theorien abweichenden Verhaltens I: "Klassische" Ansätze. Eine Einführung für Soziologen, Psychologen, Juristen, Journalisten und Sozialarbeiter. 8., überarbeitete Auflage* ([Verschiedene Aufl.]). Paderborn: W. Fink.
- Liesen, C. & Luder, R. (2011). Forschungsstand zur integrativen und separativen schulischen Förderung von Schülerinnen und Schülern mit Verhaltensauffälligkeiten. *Schweizerische Zeitschrift für Heilpädagogik*, 17(8), 11-18.
- Lipowsky, F. (2002). Zur Qualität offener Lernsituationen im Spiegel empirischer Forschung - auf die Mikroebene kommt es an. In U. Drews & W. Wallrabenstein (Hrsg.), *Freiarbeit in der Grundschule. Offener Unterricht in Theorie, Forschung und Praxis* (S. 126-159). Frankfurt, Main: Grundschulverband.
- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 73-101). Heidelberg: Springer.
- Little, E. (2005). Secondary school teachers' perceptions of students' problem behaviours. *Educational Psychology*, 25(4), 369–377.
- Lohmann, G. (2003). *Mit Schülern klarkommen. Professioneller Umgang mit Unterrichtsstörungen und Disziplin Konflikten*. Berlin: Cornelsen.
- Makarova, E., Herzog, W. & Schönbächler, M.-T. (2014). Wahrnehmung und Interpretation von Unterrichtsstörungen aus Schülerperspektive sowie aus Sicht der Lehrpersonen [Perception and interpretation of classroom disturbances in the perspectives of students and teachers]. *Psychologie in Erziehung und Unterricht*, 61(2), 127–140.
- Makarova, E., Schönbächler, M.-T. & Herzog, W. (2008). Klassenmanagement und kulturelle Heterogenität: Projektphase 1. Forschungsbericht Nr. 33.

- Makarova, E., Schönbächler, M.-T. & Herzog, W. (2009). Klassenmanagement und kulturelle Heterogenität: Ergebnisse 1. Forschungsbericht Nr. 35. Universität Bern.
- Mandel, D. (in Vorbereitung). *Schulische Integration ist (nicht) machbar. Einstellungen von Zürcher Klassenlehrpersonen zu den Einflussfaktoren auf die Realisierbarkeit*. Dissertation, Universität Zürich, Zürich.
- Marusic-Würscher, C. & Reusser, K. (2018). „Störenfriede raus“ – Zum Umgang der Lehrpersonen mit Unterrichtsstörungen (Poster), *SGBF Kongress. Bildung – Politik – Staat*. Zürich.
- Mayr, J. (2006). Klassenführung auf der Sekundarstufe II: Strategien und Muster erfolgreichen Lehrerhandelns. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(2), 227-242.
- Meijer, C. J. W. (2003). *Inclusive Education and Classroom Practice. Summary Report*. Odense: European Agency for Development in Special Needs Education.
- Menzel, D. (2009). *Verhaltensauffällige Schüler Symptome, Ursachen und Handlungsmöglichkeiten*. Bad Heilbrunn: Klinkhardt.
- Merton, R. K. (1949/1995). *Soziologische Theorie und soziale Struktur*. Berlin: de Gruyter.
- Möller, J. & Trautwein, U. (2009). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 180-203). Heidelberg: Springer.
- Montuoro, P. & Lewis, R. (2015). Student perceptions of misbehavior and classroom management. In E. T. Emmer & E. J. Sabornie (Hrsg.), *Handbook of classroom management* (2nd ed., S. 344–362). New York, NY: Routledge.
- Moser Opitz, E. (2014). Inklusive Didaktik im Spannungsfeld von gemeinsamem Lernen und effektiver Förderung. Ein Forschungsüberblick und eine Analyse von didaktischen Konzeptionen für inklusiven Unterricht. In K. Zierer (Hrsg.), *Jahrbuch für Allgemeine Didaktik. Thementeil Allgemeine Didaktik für eine inklusive Schule* (S. 52-68). Baltmannsweiler: Schneider.
- Müller, C. M. (2010). Negative peer influence in special needs classes - a risk for students with problem behaviour? *European journal of special needs education*, 25, 431-444.
- Müller, C. M. (2011). Mechanismen negativer Beeinflussung zwischen Jugendlichen mit dissozialem Verhalten und ihre Bedeutung für schulische Präventionsansätze. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*(4), 297-309.
- Müller, C. M. (2013). Negativen Peereinfluss auf Verhaltensprobleme vermeiden – Was kann die Schule tun? *Zeitschrift für Heilpädagogik*(11), 452-460.
- Müller, C. M., Fleischli, J. & Hofmann, V. (2013). *Verhaltensprobleme von Jugendlichen auf der Sekundarstufe I*: Universität Freiburg.
- Müller, C. M. & Hofmann, V. (2016). Does being assigned to a low school track negatively affect psychological adjustment? A longitudinal study in the first year of secondary school. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 27(2), 95–115.
- Müller, C. M., Hofmann, V. & Studer, F. (2012). Lässt sich individuelles Problemverhalten durch das Niveau an Verhaltensschwierigkeiten unter den Mitschülern vorhersagen? Ergebnisse einer Querschnittstudie und ihre Relevanz für die Frage einer integrativen vs. separativen Beschulung verhaltensauffälliger Schüler. *Empirische Sonderpädagogik*, 4(2), 111-128.
- Munn, P., Sharp, S., Lloyd, G., Macleod, G., McCluskey, G., Brown, J. et al. (2013). A comparison of staff perceptions of behaviour in Scottish schools in 2009 and 2006. *Research Papers in Education*, 28(2), 135–154.

- Muthén, B. O. & Muthén, L. K. (2017). *Mplus user's guide* (8th). Los Angeles, CA: Muthén & Muthén.
- Myschker, N. & Stein, R. (2014). *Verhaltensstörungen bei Kindern und Jugendlichen. Erscheinungsformen - Ursachen - hilfreiche Massnahmen*. Stuttgart: Kohlhammer.
- Nickel, H. (1985). Die Lehrer-Schüler-Beziehung aus der Sicht neuerer Forschungsergebnisse. In R. Bierman (Hrsg.), *Interaktion-Unterricht-Schule* (S. 254–280). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Niggli, A. (2013). *Didaktische Inszenierung binnendifferenzierter Lernumgebungen. Theorie - Empirie - Konzepte - Praxis*. Bad Heilbrunn: Klinkhardt.
- Nolting, H.-P. (2012). *Störungen in der Schulklasse* (10. Aufl.). Weinheim/ Basel: Beltz Verlag.
- Nunner-Winkler, G. (2012). Moral. In W. Schneider & U. Lindenberg (Hrsg.), *Entwicklungspsychologie* (S. 521-541). Weinheim: Beltz.
- Ohan, J., Visser, T. A. W., Strain, M. C. & Allen, L. (2011). Teachers' and education students' perceptions of and reactions to children with and without the diagnostic label "ADHD". *Journal of School Psychology*, 49, 81–105.
- Ophardt, D. & Thiel, F. (2013). *Klassenmanagement. Ein Handbuch für Studium und Praxis*. Stuttgart: Kohlhammer.
- Opp, G. (2009). Gefühls- und Verhaltensstörungen. Begriffsdiskussion, Erscheinungsformen, Prävalenz. In G. Opp, G. Theunissen & J. Teichmann (Hrsg.), *Handbuch schulische Sonderpädagogik* (S. 227-231). Bad Heilbrunn: Klinkhardt.
- Pappa, I. (2012). *Chancen und Herausforderungen im Umgang mit Heterogenität aus der Sicht von Lehrpersonen. Ergebnisse aus der Online-Daten-Befragung von Lehrpersonen zum Umgang mit Heterogenität nach der Einführung des neuen Volksschulgesetzes im Kanton Zürich. Unveröffentlichte Masterarbeit*. Universität Zürich.
- Pastega, N. (2015, 18. Oktober). Sonderschüler treiben Lehrer ans Limit. *SonntagsZeitung*.
- Pauli, C. & Reusser, K. (2000). Zur Rolle der Lehrperson beim kooperativen Lernen. *Schweizerische Zeitschrift für Bildungswissenschaften*, 22(3), 421-442.
- Pauli, C., Reusser, K. & Grob, U. (2010). Reformorientierter Mathematikunterricht in der Deutschschweiz. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität* (S. 309-340). Münster: Waxmann.
- Petermann, F. & Natzke, H. (2008). Aggressives Verhalten in der Schule. Ausdrucksformen, Verlaufsmuster und Möglichkeiten entwicklungsorientierter Prävention. *Zeitschrift für Pädagogik*, 54(4), 532-554.
- Peterssen, W. H. (1982). *Handbuch Unterrichtsplanung Grundfragen, Modelle, Stufen, Dimensionen*. München: Ehrenwirth.
- Pfitzner, M. & Schoppek, W. (2000). Gemeinsamkeiten und Diskrepanzen in der Bewertung von Unterrichtsstörungen durch Lehrer und Schüler. *Unterrichtswissenschaft*, 28(4), 350–378.
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft*, 13(1), 121-148.
- Pietsch, M. (2013). Was guten Unterricht kennzeichnet. *Bildung & Wissenschaft*, 12, 24-27.
- Poscher, R., Langer, T. & Rux, J. (2008). *Gutachten zu den völkerrechtlichen und innerstaatlichen Verpflichtungen aus dem Recht auf Bildung nach Art. 24 des UN-Abkommens über die Rechte von Menschen mit Behinderungen und zur Vereinbarkeit des deutschen Schulrechts mit den Vorgaben des Übereinkommens (Korrigierte Online Fassung vom 16. September 2008)* (PDF).

- Prengel, A. (2010). Heterogenität als Theorem der Grundschulpädagogik. *Zeitschrift für Grundschulforschung*, 3(1), 7-17.
- Preuss-Lausitz, U. (2005). *Verhaltensauffällige Kinder integrieren. Zur Förderung der emotionalen und sozialen Entwicklung [Integrating children with behavioural problems. Fostering the emotional and social development]*. Weinheim: Beltz.
- Preuss-Lausitz, U. & Textor, A. (2006). Das "Emsoz-Projekt": Zur Integration schwieriger Kinder in der Grundschule. In A. Platte, S. Seitz & K. Terfloth (Hrsg.), *Inklusive Bildungsprozesse* (S. 132-139). Bad Heilbrunn: Klinkhardt.
- Prüfer, P. & Rexroth, M. (2000). Zwei-Phasen-Pretesting. ZUMA-Arbeitsbericht 2000/08. Mannheim: ZUMA.
- Rajecki, D. W. (1990). *Attitudes* (2nd). Sunderland, MA: Sinauer Associates.
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2010). *Quantitative Methoden. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. Band 1* (3. Aufl.). Berlin: Springer.
- Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A. & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *The Journal of Special Education*, 38(3), 130-143.
- Reusser, K. (1999). Und sie bewegt sich doch. *Aber man behalte die Richtung im Auge. Zum Wandel der Schule und zum neu-alten pädagogischen Rollenverständnis von Lehrerinnen und Lehrern. Die neue Schulpraxis, Themenheft*, 99, 11-15.
- Reusser, K. (2008). Empirisch fundierte Didaktik - didaktisch fundierte Unterrichtsforschung. Eine Perspektive zur Neuorientierung der Allgemeinen Didaktik. In *Perspektiven der Didaktik*. (S. S. 219-237). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Reusser, K. (2011). Unterricht und Klassenführung. In L. Criblez, B. Müller & J. Oelkers (Hrsg.), *Die Volksschule zwischen Innovationsdruck und Reformkritik* (S. 68-83). Zürich: Neue Zürcher Zeitung.
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität - Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität* (S. 9-32). Münster: Waxmann.
- Reusser, K., Pauli, C. & Elmer, A. (2011). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 478-495). Münster: Waxmann.
- Reusser, K., Stebler, R., Mandel, D. & Eckstein, B. (2013). *Erfolgreicher Unterricht in heterogenen Lerngruppen auf der Volksschulstufe des Kantons Zürich. Wissenschaftlicher Bericht*: Bildungsdirektion des Kantons Zürich.
- Rosemann, B. & Bielski, S. (2001). *Einführung in die pädagogische Psychologie*. Weinheim: Beltz.
- Ryser, M. & von Erlach, E. (2007). *Bildungs mosaik Schweiz. Bildungsindikatoren 2007*. Neuchâtel: Bundesamt für Statistik (BFS).
- Schermelleh-Engel, K. & Schweizer, K. (2008). Multitrait-Multimethod-Analysen. In *Testtheorie und Fragebogenkonstruktion* (S. 325-341): Springer.
- Scherzinger, M., Wettstein, A. & Wyler, S. (2017). Unterrichtsstörungen aus der Sicht von Schülerinnen und Schülern und ihren Lehrpersonen. Ergebnisse einer Interviewstudie zum subjektiven Erleben von Störungen [Classroom disturbances in the perspectives of students and teachers. Findings of an interview study on the subjective experience of

- disturbances] *Vierteljahresschrift für Heilpädagogik und ihre Nachbarggebiete*, 86(1), 70–83.
- Schönbächler, M.-T., Makarova, E., Herzog, W., Altin, Ö., Känel, S., Lehmann, V. et al. (2009). *Klassenmanagement und kulturelle Heterogenität: Ergebnisse 2. Forschungsbericht Nr. 37. [Classroom management and cultural heterogeneity: Results 2. Research report no. 37]*. Bern: Universität Bern.
- Schwab, S. (2015). Wollen wir die verhaltensauffälligen Schüler wirklich integrieren? Die Einstellung von Österreichern über die schulische Integration von Kindern mit Verhaltensauffälligkeiten. In A. Leonhardt, K. Müller & T. Truckenbrodt (Hrsg.), *Die UN-Behindertenrechtskonvention und ihre Umsetzung. Beiträge zur Interkulturellen und International vergleichenden Heil- und Sonderpädagogik* (S. 249-257). Bad Heilbrunn: Klinkhardt.
- Schweer, M. & Thies, B. (2000). Situationswahrnehmung und interpersonales Verhalten im Klassenzimmer. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion* (S. 59–78). Opladen: Leske & Budrich.
- Schwenck, C. (2016). Psychische Störungen in der Schule. In K. Seifried, S. Drewes & M. Hasselhorn (Hrsg.), *Handbuch Schulpsychologie. Psychologie für die Schule* (2., vollständig überarb. Aufl. ed., S. 231-240). Stuttgart: Kohlhammer.
- Seidel, T. (2009). Klassenführung. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 136-148). Heidelberg: Springer Medizin-Verl.
- Seitz, W. & Stein, R. (2010). Verhaltensstörungen. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 919-927). Weinheim/Basel: Beltz Verlag.
- Sherman, J., Rasmussen, C. & Baydala, L. (2008). The impact of teacher factors on achievement and behavioural outcomes of children with Attention Deficit/Hyperactivity Disorder (ADHD): A review of the literature. *Educational Research*, 50(4), 347–360.
- Smale, G. G. (1980). *Die sich selbst erfüllende Prophezeiung. Positive oder negative Erwartungshaltungen und ihre Auswirkungen auf die pädagogische und therapeutische Beziehung*. Freiburg im Breisgau: Lambertus.
- Spiewak, M. (2010, 5. November). Verhaltensauffällige Kinder: „Die Not ist riesengroß“. *Die Zeit*.
- Spiewak, M. (2014, 5. Juni). Du störst! *Die Zeit*.
- Stallberg, F. W. & Springer, W. (1983). *Soziale Probleme. Grundlegende Beiträge zu ihrer Theorie und Analyse*. Neuwied <etc.>: Luchterhand.
- Stechow, E. v. (2015). *Von Störern, Zerstreuten und ADHS-Kindern. Eine Analyse historischer Sichtweisen und Diskurse auf die Bedeutung von Ruhe und Aufmerksamkeit im Unterricht vom 16. bis zum 21. Jahrhundert*. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Stein, R. (2014). Moralisches Urteilen und Handeln. In F. B. Wember (Hrsg.), *Handlexikon Lernschwierigkeiten und Verhaltensstörungen* (S. 154-157). Stuttgart: Kohlhammer.
- Stein, R. (2015). *Grundwissen Verhaltensstörungen* (Bd. 4., neu überarbeitete Auflage). Baltmannsweiler: Schneider Verlag Hohengehren GmbH.
- Stein, R. & Stein, A. (2014). *Unterricht bei Verhaltensstörungen*. Bad Heilbrunn: Klinkhardt.
- Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics (6th ed.)*. Boston, MA: Pearson Education.
- Textor, A. (2007). *Analyse des Unterrichts mit "schwierigen" Kindern. Hintergründe, Untersuchungsergebnisse, Empfehlungen*. Bad Heilbrunn: Klinkhardt.

- Textor, A. (2009). Offener Unterricht in der Grundschule mit Schülern mit dem Förderschwerpunkt emotionale und soziale Entwicklung. In C. Röhner, C. Henrichwark & M. Hopf (Hrsg.), *Europäisierung der Bildung* (S. 281-285). Wiesbaden: VS.
- Tillmann, K.-J. (2008). Die homogene Lerngruppe - oder: System jagd Fiktion. In H.-U. Otto & T. Rauschenbach (Hrsg.), (S. 33-39). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd). Alexandria, VA: ASCD.
- Trautmann, M. & Wischer, B. (2008). Das Konzept der Inneren Differenzierung - eine vergleichende Analyse der Diskussion der 1970er Jahre mit dem aktuellen Heterogenitätsdiskurs. *Zeitschrift für Erziehungswissenschaft*, 10(9), 159-172.
- Trautmann, M. & Wischer, B. (2011). *Heterogenität in der Schule. Eine kritische Einführung. 1. Aufl.* Wiesbaden: VS Verl. für Sozialwiss.
- Tuvblad, C., Zheng, M., Raine, A. & Baker, L. A. (2009). A common genetic factor explains the covariation among ADHD ODD and CD symptoms in 9–10 year old boys and girls. *Journal of Abnormal Child Psychology*, 37(2), 153–167.
- Ulich, K. (1980). Zur aktuellen Diskussion über abweichendes Schülerverhalten (Einführung). In K. Ulich (Hrsg.), *Wenn Schüler stören. Analyse und Therapie abweichenden Schülerverhaltens* (S. 9-16). München: Urban und Schwarzenberg.
- UNESCO. (1994). Die Salamanca Erklärung und der Aktionsrahmen zur Pädagogik für besondere Bedürfnisse. Angenommen von der Weltkonferenz "Pädagogik für besondere Bedürfnisse: Zugang und Qualität". Salamanca, Spanien, 7.-10. Juni 1994.
- Wachtel, P. (2010). Situation und Perspektiven des Förderschwerpunkts Emotionale und Soziale Entwicklung. In H. Ricking (Hrsg.), *Förderbedarf in der emotionalen und sozialen Entwicklung Prävention, Interdisziplinarität und Professionalisierung* (S. 14-28). Bad Heilbrunn: Klinkhardt.
- Wagner, W., Gollner, R., Werth, S., Voss, T., Schmitz, B. & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, 108(5), 705–721.
- Wannack, E. & Herger, K. (2014). *Classroom Management : Unterrichtsgestaltung in der Schuleingangsstufe*. Bern: hep.
- Weinert, F. E. (1997). Notwendige Methodenvielfalt. Unterschiedliche Lernfähigkeiten erfordern variable Unterrichtsmethoden. *Friedrich-Jahresheft*(XV: Lernmethoden, Lehrmethoden), 50-52.
- Wettstein, A. (2008). *BASYS. Beobachtungssystem zur Analyse aggressiven Verhaltens in schulischen Settings. Kategorienheft*. Bern: Verlag Hans Huber.
- Wettstein, A. (2012). A conceptual frame model for the analysis of aggression in social interactions. *Journal of Social, Evolutionary, and Cultural Psychology*, 6(2), 141–157.
- Wettstein, A. (2013). Die Wahrnehmung sozialer Prozesse im Unterricht. *Schweizerische Zeitschrift für Heilpädagogik*, 7(8), 5–13.
- Wettstein, A., Ramseier, E., Scherzinger, M. & Gasser, L. (2016). Unterrichtsstörungen aus Lehrer- und Schülersicht. Aggressive und nicht aggressive Störungen im Unterricht aus der Sicht der Klassen-, einer Fachlehrperson und der Schülerinnen und Schüler [Classroom disturbances in the perspectives of teachers and students. Aggressive and non-aggressive disturbances during instruction in the view of class teachers, subject teachers, and students]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48(4), 171–183.

- Wettstein, A., Scherzinger, M. & Ramseier, E. (2018). Unterrichtsstörungen, Beziehung und Klassenführung aus Lehrer-, Schüler- und Beobachterperspektive [Classroom disturbances, relationship, and classroom management in the perspectives of teachers, students and external observers]. *Psychologie in Erziehung und Unterricht*, 65(1), 58-74.
- Wettstein, A., Thommen, B. & Eggert, A. (2010). Die Bedeutung didaktischer Aspekte in der Aggressionsprävention – drei Videostudien. *Psychologie in Erziehung und Unterricht*, 57(2), 88–106.
- Wicki, W. & Kappeler, S. (2007). *Beobachtete Unterrichtsstörungen bei erfahrenen Lehrpersonen im Spiegel subjektiver Ursachenzuschreibungen*. Unveröffentlichtes Manuskript, Luzern.
- Willmann, M. (2010). Schulische Erziehungshilfe. In B. Ahrbeck & M. Willmann (Hrsg.), *Pädagogik bei Verhaltensstörungen. Ein Handbuch* (S. 67-75). Stuttgart: Kohlhammer.
- Winkel, R. (2011). *Der gestörte Unterricht. Diagnostische und therapeutische Möglichkeiten* (10.). Baltmannsweiler: Schneider.
- Wolfisberg, C. (2008). Der institutionelle Umgang mit der Heterogenität der Schulkinder. In D. Tröhler (Hrsg.), *Zukunft bilden. Die Geschichte der modernen Zürcher Volksschule*. (S. 189-199). Zürich: Verlag Neue Zürcher Zeitung NZZ Libro.
- Zevenbergen, R. (2001). Mathematics, social class, and linguistic capital. In B. Atweh, H. Forgasz & B. Nebres (Hrsg.), *Sociocultural research on mathematics education* (S. 201–215). Mahwah, NJ: Erlbaum.

5. Tabellen- und Abbildungsverzeichnis

Tabelle 1. Unterrichtliche Devianz – Übersicht über alle Items.....	99
Tabelle 2. Subjektives Störungsempfinden – Übersicht über alle Items.....	100
Abbildung 1. Prädiktion von Integrationsoptimismus im SEM (Eckstein et al., 2013, S. 12).	23
Abbildung 2. Interaktionistisches Theoriemodell der Produktion und Rezeption gestörten Unterrichts.	60
Abbildung 3. Mehrebenenmodell der Peer-Ratings (Eckstein, Grob, et al., 2016, S. 123).	66
Abbildung 4. Deskriptiver Mittelwertsvergleich nach Rater- und Zielschülergruppe.	69
Abbildung 5. Perspektivenvergleich: Schüler/innen vs. Lehrperson innerhalb der Beispielklasse.	72
Abbildung 6. Zielkindspezifische Auftretenshäufigkeit unterrichtlicher Devianz gemäss den Lehrpersonen.	84
Abbildung 7. Mittleres zielschülerspezifisches Störungsempfinden der Lehrpersonen.	85
Abbildung 8. Perspektivenvergleich Lp/SuS innerhalb der Durchschnittsklasse.	90
Abbildung 9. Klasseninterne Perspektivenvergleiche Lp/SuS in sechs Klassen.	92
Abbildung 10. Varianzanalytischer Perspektivenvergleich des Konstrukts undiszipliniertes Verhalten.	102
Abbildung 11. Varianzanalytischer Perspektivenvergleich des Konstrukts dissoziales Verhalten.	103
Abbildung 12. Varianzanalytischer Perspektivenvergleich des Konstrukts affektives Störungsempfinden.	104
Abbildung 13. Varianzanalytischer Perspektivenvergleich des Konstrukts kognitives Störungsempfinden.	105
Abbildung 14. Resultate des Twolevel CT-C(M-1) Modells (Eckstein, under review).	110
Abbildung 15. Skizze des ursprünglich geplanten Drei-Ebenen-Modells.	135
Abbildung 16. Ergebnisse des Pfadmodells nach neuer Analysestrategie.	138

6. Anhang

6.1. Originalarbeiten

6.1.1. Umsetzung der integrativen Volksschule – Was Lehrpersonen optimistisch macht³⁰

Abstract

Seit der Annahme des neuen Zürcher Volksschulgesetzes im Jahr 2005 ist das kantonale Bildungssystem dazu verpflichtet, möglichst alle Lernenden mit besonderen pädagogischen Bedürfnissen integrativ in Regelklassen zu unterrichten. Die Lehrpersonen sind aufgefordert, in heterogenen Lerngruppen Unterricht zu gestalten, der Schüler/-innen mit unterschiedlichsten Voraussetzungen und Bedürfnissen optimal fördert. Der vorliegende Artikel geht der Frage nach, unter welchen Voraussetzungen Lehrpersonen optimistisch sind, dass die Umsetzung der integrativen Volksschule gelingt. Die im Rahmen des Heterogenitäts-Projekts (Universität Zürich, Institut für Erziehungswissenschaft) via eine standardisierte Online-Befragung bei 385 Lehrpersonen aus 19 Schulen im Kanton Zürich erhobenen Daten werden genutzt, um in einem Strukturgleichungsmodell den Einfluss von vier Prädiktoren auf das latente Konstrukt ‚Integrationsoptimismus‘ zu analysieren. Die Ergebnisse zeigen, dass sich der Integrationsoptimismus der Lehrpersonen vor allem durch eine Erweiterte Didaktische Unterrichtspraxis sowie die Quote verhaltensauffälliger Schüler/-innen pro Klasse vorhersagen lässt. In der Diskussion werden das Auswertungsverfahren und die praktische Relevanz der Befunde reflektiert.

³⁰ Copyright © 2013 by Schweizerische Gesellschaft für Bildungsforschung SGBF. Reproduced [or adapted] with permission. The official citation that should be used in referencing this material is: Eckstein, B., Reusser, K., Stebler, R. & Mandel, D. (2013). Umsetzung der integrativen Volksschule – Was Lehrpersonen optimistisch macht: Eine Analyse der Überzeugungen von Klassenlehrpersonen im Kanton Zürich. *Schweizerische Zeitschrift für Bildungswissenschaften*, 35(1), 91-112.

Einleitung

Mit der Ausrichtung des neuen Volksschulgesetzes des Kantons Zürich (Kantonsrat des Kantons Zürich, 2005) sollen neu auch Schüler/-innen mit besonderen pädagogischen Bedürfnissen möglichst in Regelklassen und nicht mehr in Sonderklassen unterrichtet werden. Integration statt Separation lautet die Devise, welche u.a. mit sonderpädagogischen Massnahmen realisiert werden soll. Deshalb muss auch die Orientierung des Unterrichts am Durchschnittsschüler schulischen Lernkulturen weichen, welche Heterogenität als Ressource nutzen. Gefragt sind eine Pädagogik der Vielfalt, didaktische Konzepte für individualisiertes Lernen sowie entsprechende institutionelle Rahmenbedingungen (Buholzer & Kummer Wyss, 2010; Prengel, 2006). Die Lehrpersonen stehen vor der anspruchsvollen Aufgabe, die integrative Volksschule im Unterricht umzusetzen. Vor diesem Hintergrund stellt sich die Frage, was die Lehrpersonen, welche die Integration bewerkstelligen bzw. erfolgreichen Unterricht in heterogenen Lerngruppen gestalten müssen, von der integrativen Volksschule halten. Im vorliegenden Beitrag werden entsprechende Überzeugungen von Lehrpersonen untersucht, indem vier Schritte unternommen werden: Theoretische Situierung des Problemzusammenhangs (Kap. 2), Erläuterung der Methode (Kap. 3), Ergebnisdarstellung (Kap. 4) und Diskussion (Kap. 5).

Optimistische Überzeugungen von Lehrpersonen bezüglich Integration

Integrationsoptimismus, verstanden als zuversichtliche, durch positive Erwartungen bestimmte Haltung der Lehrpersonen gegenüber der Integration im Bildungswesen, dürfte die Umsetzung der integrativen Schulform des neuen Volksschulgesetzes im Kanton Zürich wesentlich beeinflussen. In der Literatur wird die Relevanz solcher Überzeugungen kontrovers

diskutiert (Altrichter, Trautmann, Wischer, Sommerau & Doppler, 2009; Trautmann & Wischer, 2011). Relativ gut belegt ist ein bedeutsamer Zusammenhang zwischen den berufsbezogenen Überzeugungen und der Qualität des professionellen Handelns von Lehrpersonen (z.B. Voss, Kleickmann, Kunter & Hachfeld, 2011), wobei die Reichweite und die prädiktive Aussagekraft dieser Beziehung nicht ausreichend geklärt sind (Reusser, Pauli & Elmer, 2011). Es werden Wechselwirkungen vermutet: Die Überzeugungen der Lehrpersonen entfalten in Interaktion mit weiteren Komponenten der professionellen Kompetenz, der berufsbiographischen Prägung und situativen Bedingungen eine gewisse Handlungsrelevanz. Umgekehrt beeinflusst das Ergebnis des unterrichtlichen Handelns die Entwicklung von Überzeugungen (ebd.). Unterricht in heterogenen Lerngruppen ist folglich zugleich Ursache und Wirkung von Überzeugungen betreffend die integrative Schulform. Innerhalb dieses komplexen Ursache-Wirkungs-Systems fokussiert der vorliegende Beitrag Bedingungen handlungsnaher, berufsbezogener Überzeugungen.

Handlungsferne und handlungsnaher Überzeugungen von Lehrpersonen

Berufsbezogene Überzeugungen von Lehrpersonen sind «affektiv aufgeladene [...] Vorstellungen über das Wesen und die Natur von Lehr-Lernprozessen, Lerninhalten, die Identität und Rolle von Lernenden und Lehrenden [...] sowie den institutionellen und gesellschaftlichen Kontext von Bildung und Erziehung, welche für wahr oder wertvoll gehalten werden» (Reusser et al., 2011, S. 478). Ausgehend von diesem Begriffsverständnis wird im Folgenden zwischen handlungsfernen und handlungsnahen Überzeugungen unterschieden.

Handlungsferne berufsbezogene Überzeugungen von Lehrpersonen sind allgemeiner Natur und können Reflexionen über makrostrukturelle Systemzusammenhänge enthalten. Sie

sind dem Bewusstsein zugänglich und veränderbar, entfalten jedoch nur eine geringe Handlungswirksamkeit (ebd.). Handlungsfern ist z.B. die Überzeugung, dass Integration im Sinne eines bildungspolitischen Grundsatzes wichtig und wertvoll ist. Es hat sich in verschiedenen Forschungsarbeiten gezeigt, dass viele Lehrpersonen so denken (Gebhardt et al., 2011; Kunz, Luder & Moretti, 2010). Allerdings ist zu vermuten, dass einige Lehrpersonen die Umsetzung der integrativen Schulform in der Gesamtbilanz kritisch beurteilen, was sie bei einer standardisierten Befragung mit Fokus auf handlungsferne Überzeugungen nur eingeschränkt zum Ausdruck bringen können. Entsprechend dieser Vermutung wurde im Rahmen des Heterogenitäts-Projekts (vgl. Kap. 3.1) anhand einer standardisierten Onlinebefragung mit N=385 Zürcher Lehrpersonen ermittelt, dass rund 80% der Befragten Integration grundsätzlich für wünschbar halten. Jedoch finden nur 40% der Befragten, dass Integration unter den gegebenen Voraussetzungen überhaupt realisierbar sei (Reusser, Stebler, Mandel & Eckstein, 2013).

Daher geraten *handlungsnahe berufsbezogene Überzeugungen* von Lehrpersonen ins Blickfeld. Diese sind spezifischer Natur und enthalten Situationsprototypen mit hoch wirksamen Steuerungsfunktionen in Bezug auf das konkrete Handeln. Sie sind dem Bewusstsein nur bedingt zugänglich, entlasten dafür das Denken und ermöglichen es, schnell und effizient zu handeln. Sie werden als relativ stabil und vergleichsweise veränderungsresistent beschrieben, da sie Handlungssicherheit gewährleisten (Reusser et al., 2011). Die folgenden Ergebnisse des Heterogenitäts-Projekts illustrieren handlungsnahe Überzeugungen von Lehrpersonen bezüglich der Umsetzung der integrativen Volksschule:

- Rund 85 Prozent der Lehrpersonen finden, dass die Klassenführung anspruchsvoller und die Unterrichtsvorbereitung aufwändiger geworden sei.

- Fast die Hälfte der Lehrpersonen sind der Ansicht, dass die Unterrichtsqualität unter der integrativen Schulform leide (Reusser et al., 2013).

Diese Befunde deuten an, dass die Lehrpersonen die Unterrichtsgestaltung seit der Gesetzesrevision als herausfordernd – zuweilen auch als überfordernd – erleben. In den folgenden zwei Abschnitten werden mögliche Ursachen dieser handlungsnahen, berufsbezogenen Überzeugungen thematisiert.

Zur Bedeutung des Unterrichts für die integrationsbezogenen Überzeugung von Lehrpersonen

Wenn Lehrpersonen gute Erfahrungen mit heterogenen Schulklassen machen, dürften optimistische Überzeugungen hinsichtlich der Realisierbarkeit der integrativen Volksschule resultieren. Somit stellt sich die Frage, wie Lehrpersonen erfolgreichen Unterricht in heterogenen Lerngruppen gestalten können. Im Hinblick auf gute Leistungsergebnisse legen Forschungsergebnisse nahe, dass der Unterricht v.a. *hinreichend differenziert* werden muss (Seidel & Shavelson, 2007; Tillmann & Wischer, 2006). Mit Binnendifferenzierung werden unterschiedliche didaktische Massnahmen in Verbindung gebracht, die den Lernenden in der gleichen Unterrichtssituation unterschiedliche Bearbeitungsmöglichkeiten eröffnen sollen (Trautmann & Wischer, 2008). Zur Systematisierung dieser Massnahmen bietet sich die Denkfigur des didaktischen Dreiecks an, dessen Kanten das Entscheidungs- und Handlungsfeld von Lehrpersonen symbolisieren (Reusser, 2008): Der Unterricht kann mit Bezug auf die Lernziele, Themen und Aufgaben (Ziel- und Stoffkultur), die Inszenierungsformen (Wissens- und Lernkultur) und die Lernunterstützung (Beziehungs- und Unterstützungskultur) differenziert werden (Reusser et al., 2013). Im Kern geht es bei sämtlichen Differenzierungsmassnahmen

darum, adaptive Lernangebote und dadurch optimale Voraussetzungen für individualisiertes Lernen zu schaffen.

Eine umfassende Form der Binnendifferenzierung sind die *Erweiterten Lehr-Lernformen* (ELF), wobei es sich dabei um eine Deutschschweizer Variante des reformpädagogisch geprägten Konzepts «Offener Unterricht» handelt (Crocì, Imgrüth, Landwehr & Spring, 1995). Die ELF sind ein didaktisches Prinzip, das den traditionellen Unterricht (z.B. Frontalunterricht) im Hinblick auf vier hauptsächliche Zielsetzungen erweitern will: (1) Förderung der Selbststeuerung durch die Lernenden, (2) Erhöhung der Adaptivität des Unterrichts, (3) Förderung der Lernstrategien (das Lernen lernen) und (4) hohe Eigenaktivität der Lernenden bei der Auseinandersetzung mit den Inhalten (Pauli, Reusser & Grob, 2010). Methodisch realisiert werden die ELF z.B. mit Werkstattunterricht oder mit Planarbeit.

In der Schweiz ist Unterricht nach dem Reformkonzept ELF verbreitet (Pauli, Reusser, Waldis & Grob, 2003; Reusser & Stebler, 2000). In Deutschland erfreut sich Offener Unterricht zunehmender Beliebtheit, wobei die Lektionen primär in organisatorischer Hinsicht geöffnet werden, Lernwege oder Inhalte hingegen meistens festgelegt sind (Hartinger, 2005). Aufgrund der qualitativ unterschiedlichen Realisierung dieser Reformkonzepte zeichnen Wirksamkeitsstudien kein einheitliches Bild (Altrichter et al., 2009; Bohl & Kucharz, 2010; Klieme & Warwas, 2011; Lipowsky, 2002; Pauli et al., 2003). Aus der Befundlage lässt sich jedoch insgesamt schliessen, dass sich die ELF für erfolgreichen Unterricht mit heterogenen Lerngruppen eignen. Diese Einschätzung teilt die Mehrheit der im Heterogenitäts-Projekt befragten Lehrpersonen: Mindestens je 60 Prozent erachten die ELF als eher geeignet oder als sehr geeignet für den Unterricht in heterogenen Lerngruppen, obwohl sie nicht alle dieser Unterrichtsformen regelmässig einsetzen (Reusser et al., 2013).

Die Forschung verdeutlicht auch, dass es weniger auf die Oberfläche des Unterrichts (u.a. Methodenspektrum, Vielfalt des Lernangebots, zeitliche und inhaltliche Freiheitsgrade) als auf die Tiefenstruktur, d.h. die kognitive Aktivierung der einzelnen Schüler/-innen, die Fokussierung auf gründliches Verstehen und die adaptive Unterstützung durch die Lehrperson ankommt (Reusser, 2011). Es zeigt sich, dass individualisierter, differenzierter oder offener Unterricht nicht per se wirksam ist, sondern mit Blick auf hohe Fachleistungen und auf leistungsschwächere Schüler/-innen intelligent organisiert sein muss (Bohl, T., Batzel & Richey, 2011; Joller-Graf, 2010).

Zur Bedeutung struktureller Rahmenbedingungen für die integrationsbezogenen Überzeugung von Lehrpersonen

Werden die Lehrpersonen direkt auf strukturelle Rahmenbedingungen angesprochen, nennen sie häufig ähnliche Gelingensbedingungen einer erfolgreichen Umsetzung der integrativen Volksschule, z.B. geeignete Räumlichkeiten; genügend bezahlte Zeitgefässe für eine produktive Kooperationskultur; mehrjähriges Bestreben nach Schul- und Unterrichtsentwicklung (Reusser et al., 2013). Zudem verweisen die Lehrpersonen ebenfalls vielfach übereinstimmend auf Risikofaktoren, z.B. umständliche Dienstwege; zu wenig Lektionen für integrative Förderung; ungenügend differenzierte Lehrmittel (ebd.). Solche Rahmenbedingungen werden gemäss dem Mehrebenen-Angebots-Nutzungs-Modell von Reusser und Pauli (2010) als Prädiktoren der berufsbezogenen Überzeugungen von Lehrpersonen verstanden, wovon in den folgenden zwei Abschnitten eine Auswahl etwas genauer vorgestellt wird.

Berufsbiographie und berufliche Funktion von Lehrpersonen

Die Berufslaufbahn der Lehrpersonen hat einen entscheidenden Einfluss auf die Entwicklung ihrer berufsbezogenen Überzeugungen (Reusser et al., 2011). Ein guter Indikator dafür ist möglicherweise das Dienstalter der Lehrpersonen. Es könnte angenommen werden, dass ein hohes Dienstalter – welches auf eine Berufsausbildung und auf Unterrichtserfahrungen unter den Vorzeichen des ehemals separativen Schulsystems verweist – die Entwicklung integrationsbezogener Überzeugungen ungünstig beeinflusst.

Weiter ist zu berücksichtigen, dass Lehrpersonen eine heterogene Berufsgruppe sind. Es ist anzunehmen, dass verschiedene *berufliche Funktionen* mit unterschiedlichen integrationsbezogenen Überzeugungen einhergehen. Regellehrpersonen dürften andere Überzeugungen haben als Speziallehrpersonen, die sich für schulische Heilpädagogik qualifiziert haben. Allerdings befinden Gebhardt et al. (2011), dass Erfahrungen mit der integrativen Schulform wichtiger sind für die Entwicklung solcher Überzeugungen als die Ausbildung.

Ein anderer relevanter Faktor ist die *Schulstufe*. Lehrpersonen der Primarstufe haben eine andere Ausbildung und andere Rahmenbedingungen als Lehrpersonen der Sekundarstufe. Wie sich im Heterogenitäts-Projekt zeigte, sehen die Lehrpersonen der Primarstufe in der integrativen Schulung mehr Potenzial als die Lehrpersonen der Sekundarstufe (Reusser et al., 2013).

Schliesslich sind *Klassenlehrpersonen* von Lehrkräften zu unterscheiden, welche keine Gesamtverantwortung für eine Klasse tragen (Regierungsrat des Kantons Zürich, 2006). Sie haben dafür zu sorgen, dass aus der heterogenen Schülerschaft eine funktionierende Lerngruppe entsteht, in der alle Schüler/-innen von den Lernangeboten optimal profitieren.

Grösse und Zusammensetzung von Lerngruppen

In *kleinen Klassen* sollte es den Lehrpersonen eher möglich sein als in grossen Klassen, individuelle Lernstände der Schüler/-innen verlässlich zu diagnostizieren und adaptive Unterrichtsangebote zu schaffen. Dadurch könnten bei den Lernenden bessere Leistungen und bei den Lehrpersonen optimistischere Einstellungen zur integrativen Schulform entstehen. Diese Vermutungen werden von der empirischen Befundlage jedoch nicht ausnahmslos gestützt (Altrichter et al., 2009; Gröhlich, Scharenberg & Bos, 2009), sodass eine detaillierte Analyse der Unterrichtsprozesse bei jeder Klassengrösse angezeigt ist (Schukajlow & Blum, 2011). Aber auch neuere Forschungsarbeiten, welche dies berücksichtigen, liefern uneinheitliche Befunde. So bescheinigen einzelne Studien kleinen gegenüber grossen Klassen positive Effekte, z.B. zeitlich mehr Aufmerksamkeit pro Schüler/-in durch die Lehrperson; ausgeprägteres «on task»-Verhalten der Lernenden (Blatchford, Bassett & Brown, 2011; Harfitt, 2012). Aus einer anderen Studie geht jedoch hervor, dass in einem stark binnendifferenzierten Unterricht unabhängig von der Klassengrösse effektive Lernprozesse ausgelöst werden (Brühwiler & Blatchford, 2011).

Des Weiteren stellt sich die Frage, inwiefern unterschiedliche Voraussetzungen und Bedürfnisse der Lernenden die Unterrichtsabläufe und dadurch vermittelt die Überzeugungen der Lehrpersonen beeinflussen. Gebhardt et al. (2011) zeigten, dass die Einstellungen der Lehrpersonen zur schulischen Integration behinderter Kinder in signifikanter Weise von der Art der Behinderung abhängt: Während die Integration körperlich behinderter Schüler/-innen durchschnittlich sehr positiv beurteilt wird, sind die Einstellungen zur Integration geistig behinderter Schüler/-innen vergleichsweise ungünstig. Des Weiteren deuten Resultate einer Voruntersuchung derselben Forschergruppe an, «dass die Einstellungen der LehrerInnen zur Integration von SchülerInnen mit *Verhaltensauffälligkeiten* [...] noch ungünstiger sind als jene

zur Integration von SchülerInnen mit einer geistigen Behinderung» (ebd., S. 287). Diese Befunde entsprechen den Ergebnissen zahlreicher weiterer Forschungsarbeiten, wonach Kinder und Jugendliche mit Verhaltensauffälligkeiten am schwierigsten in Regelklassen zu integrieren sind (Liesen & Luder, 2011; Preuss-Lausitz & Textor, 2006). Entsprechend fordern viele Lehrpersonen mehr Unterstützung für ihre Arbeit mit verhaltensauffälligen Schüler/-innen (Reusser et al., 2013).

Fazit und Fragestellung

Die referierten Befunde lassen sich wie folgt zusammenfassen: Die Lehrpersonen sind mehrheitlich überzeugt, dass Integration wichtig und wünschbar ist (handlungsferne Überzeugungen). Sie sind jedoch eher skeptisch hinsichtlich der Implementation der integrativen Schulform (handlungsnahe Überzeugungen). Der vorliegende Artikel fokussiert diese handlungsnahen Überzeugungen und erörtert Voraussetzungen für den Integrationsoptimismus, also für die Zuversicht der Lehrpersonen betreffend die Realisierbarkeit der integrativen Volksschule. Es wird zum einen angenommen, dass Lehrpersonen, welche qualitativ hochwertigen binnendifferenzierten Unterricht gestalten, einen vergleichsweise erfolgreichen Umgang mit heterogenen Lerngruppen erleben und folglich tendenziell zuversichtlich sind, dass die integrative Volksschule realisierbar ist. Zum anderen wird angenommen, dass strukturelle Rahmenbedingungen (z.B. Berufsbiographie, Zusammensetzung der Lerngruppe) die integrationsbezogenen Überzeugungen der Lehrpersonen beeinflussen. Davon ausgehend wird im empirischen Teil des vorliegenden Artikels folgende Fragestellung untersucht:

Welchen Einfluss haben unterrichtsbezogene und strukturelle Faktoren auf den Optimismus von Lehrpersonen bezüglich der Umsetzung der integrativen Volksschule?

Methode

Studiendesign und Analysestichprobe

Der vorliegende Artikel basiert auf Daten des *Heterogenitäts-Projekts* (Laufzeit: 2010-2013). Es handelt sich hierbei um eine Auftragsstudie der Bildungsdirektion Zürich an den Lehrstuhl ‚Pädagogische Psychologie und Didaktik‘ am Institut für Erziehungswissenschaft der Universität Zürich, in welcher verschiedene Methoden angewandt wurden (u.a. Leitfadeninterviews mit Schulleitungspersonen, standardisierte Onlinebefragung von Lehrpersonen). Die ersten Datenauswertungen für den wissenschaftlichen Bericht lieferten vorwiegend deskriptive Ergebnisse – beispielsweise den genannten Befund bezüglich des tendenziell gering ausgeprägten Integrationsoptimismus der Zürcher Lehrpersonen (Reusser et al., 2013). Nachfolgend werden weiterführende Analysen zu diesem Befund präsentiert, die sich auf den Datensatz der Onlinebefragung stützen.

Die Stichprobenziehung erfolgte in Zusammenarbeit mit der Auftraggeberin (BI ZH) nach einem bewussten Auswahlverfahren: Insgesamt N=19 Zürcher Volksschulen wurden ausgewählt, um sowohl Schulen im Sample vertreten zu haben, welche bereits einen gelingenden Unterricht mit heterogenen Lerngruppen praktizieren als auch solche, welche diesbezüglich noch Entwicklungsbedarf aufweisen. Von den N=385 befragten Lehrpersonen der Gesamtstichprobe wurden im Hinblick auf möglichst ähnliche, strukturelle Voraussetzungen nur die Klassenlehrpersonen des Regelschulbereichs der Primarstufe (n=92) sowie der Sekundarstufe (n=93) berücksichtigt. Die Analysestichprobe zählt somit total N=185 Lehrpersonen.

Skizze des Strukturgleichungsmodells und Beschreibung der verwendeten Items

Das weiter unten detailliert dargestellte Strukturgleichungsmodell (SEM) bildet das zentrale Element des vorliegenden Beitrags. Bei diesem Schätzverfahren handelt es sich grundsätzlich um eine Kausalanalyse, in welcher der Einfluss unabhängiger Variablen (Prädiktoren) auf die abhängige Variable (Kriterium) ‚Integrationsoptimismus‘ geschätzt wird. Diese Variablen werden in den folgenden Abschnitten eingeführt, beginnend mit dem Kriterium ‚Integrationsoptimismus‘. Anschliessend werden die Prädiktoren behandelt, welche mit einem stark unterrichtsbezogenen Fokus ausgewählt wurden.

Der hauptsächliche Grund für diese Auswahl besteht darin, dass der Unterricht als Kerngeschäft der Volksschule eine zentrale Referenzgrösse darstellt: Ist der Unterricht erfolgreich (z.B. gemessen an motivationalen und leistungsbezogenen Outcomes bei den Lernenden), kann das Bildungssystem insgesamt als erfolgreich bezeichnet werden. Sodann ist davon auszugehen, dass auch die Einstellungen und Überzeugungen der Lehrpersonen mit Merkmalen der Gestaltung und mit der Qualität des Unterrichts zusammenhängen. Allerdings ist Unterrichtsqualität immer auch bedingt durch übergeordnete Systemzusammenhänge. Insofern wäre es ebenfalls interessant, den Einfluss struktureller Rahmenbedingungen auf den Integrationsoptimismus der Lehrpersonen zu schätzen. Da in einem Strukturgleichungsmodell jedoch eine Begrenzung der zu berechnenden Informationen erforderlich ist, werden solche strukturellen Rahmenbedingungen zugunsten «unterrichtsnäherer» Variablen etwas weniger fokussiert, was sich in der Auswahl der Prädiktoren widerspiegelt.

Integrationsoptimismus

Die Zuversicht der Lehrpersonen betreffend die Realisierbarkeit der integrativen Volksschule wurde im Online-Fragebogen als theoretisches Konstrukt erfasst. Der resultierenden Variable ‚Integrationsoptimismus‘ liegen sechs Items zugrunde, die auf einer sechsstufigen Skala beantwortet wurden (von 1 = «Ich stimme überhaupt nicht zu» bis 6 = «Ich stimme voll und ganz zu»). Z.B. «Die Integration von Schüler/-innen (SuS) mit besonderen pädagogischen Bedürfnissen ist eine Bereicherung für die Regelklasse». Die Formulierung der Items war darauf ausgelegt, relativ handlungsnahen Überzeugungen von Lehrpersonen bezüglich Integration zu erfassen. Im Strukturgleichungsmodell stellt ‚Integrationsoptimismus‘ die Kriteriumsvariable dar und wird als latentes Konstrukt anhand von drei Parcels des Typs ‚Mittelwertscore‘ aus jeweils zwei Items gebildet¹. Die Eindimensionalität des Konstrukts wurde anhand explorativer Faktorenanalysen mit den manifesten Variablen ermittelt². Die Paarung der Items zur Bildung der Mittelwertcores der Parcels erfolgte anhand der Rangplätze der sechs Items innerhalb der Komponente, welche die explorative Faktorenanalyse ermittelte, sowie aufgrund der ursprünglichen Polung der Items, um eine möglichst gute Durchmischung im Hinblick auf das latente Konstrukt im Strukturgleichungsmodell zu erzielen.

Erweiterte Didaktische Unterrichtspraxis (EDUP)

Der Prädiktor ‚Erweiterte Didaktische Unterrichtspraxis‘ (EDUP) basiert auf insgesamt 14 manifesten Items und wurde in den SEM-Analysen als latentes Konstrukt zweiter Ordnung modelliert. Es besteht aus folgenden zwei Komponenten, die jeweils als latente Konstrukte erster Ordnung modelliert wurden.

Binnendifferenzierung (DIFF)

Die Lehrpersonen gaben bei 9 Items auf einer sechsstufigen Skala an, inwiefern sie ihren Unterricht binnendifferenziert gestalten (von 1 = «Trifft überhaupt nicht zu» bis 6 = «Trifft ganz genau zu»). Z.B. «Schnellere SuS gehen schon zum nächsten über, wenn ich mit den Langsameren noch übe oder wiederhole». Mittels explorativer Faktorenanalyse in SPSS wurde eine inhaltlich begründbare, zweidimensionale Struktur herausgearbeitet: Erstens laden vier Items auf einen gemeinsamen Faktor, der binnendifferenzierende Massnahmen mit Bezug auf die unterschiedlichen Interessen und Neigungen der Lernenden thematisiert (z.B.

«Die SuS können bei mir eigene Themen, Aufgaben, Lesetexte etc. wählen»). Zweitens laden fünf Items auf einen Faktor, der die leistungsbezogene Binnendifferenzierung thematisiert (z.B. «Die SuS erhalten von mir ihren Fähigkeiten entsprechende unterschiedliche Hausaufgaben»). Aus diesen vier bzw. fünf Items wurden zwei Mittelwertscores gebildet, die im Strukturgleichungsmodell als Parcels für die Schätzung des latenten Konstrukts DIFF eingesetzt wurden.

Erweiterte Lehr-Lernformen (ELF)

Die Lehrpersonen gaben auf einer vierstufigen Skala an, wie häufig sie im letzten halben Jahr folgende fünf Unterrichtsformen eingesetzt hatten (von 1 = «in höchstens 1 Lektion pro Monat» bis 4 = «praktisch jeden Tag»): Planarbeit, Werkstattunterricht, Freiarbeit, Arbeiten in Lernpartnerschaften, Einsatz von lernstarken SuS als Tutoren für lernschwache SuS. Gemäss explorativer Faktorenanalyse bilden diese fünf Items eine eindimensionale Struktur. Zur Schätzung des latenten Konstrukts ELF wurden wiederum Parcels aus den Mittelwertscores von zwei bzw. drei dieser ELF-Items gebildet.

Dienstalter, Klassengrösse, Quote Verhaltensauffälliger pro Klasse (VA-Quote)

Neben der Variable EDUP werden im Strukturgleichungsmodell drei weitere Variablen als Prädiktoren eingeführt. Zur Beantwortung dieser Items stand den Lehrpersonen ein Eingabefeld zur Verfügung, wo sie die entsprechenden Ziffern per Tastatur eintragen konnten:

- *Dienstalter*. Exakter Wortlaut der Frage: «Wie viele Jahre Unterrichtspraxis haben Sie insgesamt?»
- *Klassengrösse*. Exakter Wortlaut der Frage: «Wie viele Schüler/innen sind in Ihrer Klasse?»
- *VA-Quote*. Sinngemässe Formulierung der Frage: Bei wie vielen Schüler/innen Ihrer Klasse trifft folgende Kategorie zu [...] ausgeprägte Verhaltensauffälligkeit?³

Eignung der Daten für die Berechnung des Strukturgleichungsmodells

SEM-Verfahren setzen intervallskalierte, normalverteilte Daten und eine gewisse Fallzahl voraus. Intervallskaliertes Datenniveau ist bei den Variablen ‚Dienstalter‘, ‚Klassengrösse‘ und ‚VA-Quote‘ gegeben. Die latenten Konstrukte ‚Integrationsoptimismus‘, ‚ELF‘ und ‚Binnendifferenzierung‘ wurden auf der Ebene der manifesten Einzelitems auf ordinalem Skalenniveau erhoben. Da für die SEM-Analysen jeweils Parcels in Form von Mittelwertscores gebildet wurden, weisen diese Variablen ähnliche Eigenschaften wie intervallskalierte Daten auf, sodass mit einem gewissen Vorbehalt davon auszugehen ist, dass diese Voraussetzung eingehalten wird.

Zwecks Überprüfung der Normalverteilung wurden die einzelnen Variablen anhand ihrer deskriptiven Kennwerte (Schiefe, Kurtosis) untersucht. Bei den Variablen ‚Klassengrösse‘ und ‚VA-Quote‘ wurden vier Ausreisser entdeckt. Diese wurden für die SEM-Analysen als

systemdefinierte Missings behandelt, die vier Fälle jedoch im Sample belassen. Nach dem Ausschluss der vier Ausreisser liegen sämtliche Werte unter den in der Literatur empfohlenen (nicht besonders strengen) kritischen Grenzwerten (Curran, West & Finch, 1996).

Explizite Hinweise auf eine Mindest-Fallzahl finden sich in der Literatur kaum, weil diese von diversen Umständen abhängt (Bentler, 2006). Die Fallzahl der verwendeten Analysestichprobe von N=185 Lehrpersonen ist als vergleichsweise moderat anzusehen. Im Strukturgleichungsmodell wurden jedoch die Primarlehrpersonen (n=92) von den Sekundarlehrpersonen (n=93) getrennt analysiert, womit die jeweilige Fallzahl als knapp ausreichend einzustufen ist. Entsprechend wurde nach der Daumenregel, dass etwa zehn Fälle pro manifester Variable im Modell benötigt werden (Byrne, 2010), keine weiteren Variablen im Modell eingeführt.

Ergebnisse

Deskriptive Befunde und Prüfung auf Mittelwertsunterschiede

Weil in der nachfolgenden Berichterstattung der SEM-Analysen kaum auf deskriptive Befunde eingegangen wird, dient diese Übersicht als eine Art Vorbereitung, um die verwendeten Variablen etwas plastischer darzustellen. Es ist zu beachten, dass diese Analysen mit SPSS anhand der manifesten Variablen vorgenommen wurden – die Konstrukte ‚Integrationsoptimismus‘, ‚ELF‘ und ‚Binnendifferenzierung‘ wurden im Strukturgleichungsmodell jedoch als latente Factorscores, also messfehlerbereinigt geschätzt.

Integrationsoptimismus: Der Median des manifesten Mittelwertscores der sechs Items liegt bei 3.00, das arithmetische Mittel beträgt 3.13 mit einer Standardabweichung (SD) von 1.05

(sechsstufige Antwortskala). Im Vergleich zum theoretischen Skalenmittelwert (3.5) bedeutet dies, dass die Lehrpersonen eine leichte Skepsis bezüglich der Realisierbarkeit der integrativen Volksschule zum Ausdruck bringen. Bezüglich der zwei Subsamples lässt sich erwartungsgemäss festhalten, dass die Lehrpersonen der Primarstufe im Durchschnitt (3.45) etwas optimistischer sind als die Sekundarlehrkräfte (2.82); dieser Unterschied ist statistisch signifikant ($t=4.31$, $df=183$, $p<.001$).

Binnendifferenzierung: Der Median des manifesten Mittelwertscores der neun Items liegt bei 4.11 (sechsstufige Antwortskala), das arithmetische Mittel beträgt 4.08 ($SD = 0.79$). Die Lehrpersonen geben also mehrheitlich an, dass sie ihren Unterricht binnendifferenziert gestalten. Das arithmetische Mittel der Primarlehrpersonen (4.40) ist signifikant höher als dasjenige der Sekundarlehrpersonen (3.77) ($t=5.85$; $df=180$; $p<.001$).

Erweiterte Lehr-Lernformen: Der Median des manifesten Mittelwertscores der fünf Items liegt bei 2.23 (vierstufige Antwortskala), das arithmetische Mittel beträgt 2.31 ($SD = 0.63$). Das bedeutet, dass die Mehrheit der Lehrpersonen diese Unterrichtsformen seltener als «in ein paar Lektionen pro Woche» einsetzen. Das arithmetischen Mittel der Primarlehrpersonen (2.49) übertrifft erneut dasjenige der Sekundarlehrpersonen (2.14) in signifikanter Weise ($t=3.71$; $df=176$; $p<.001$).

Dienstalter: Im Durchschnitt verfügen die Lehrpersonen über 16.09 Jahre Berufserfahrung mit einer Standardabweichung von 11.61. Der Median liegt bei 13.00 Jahren, was insgesamt auf eine ziemlich symmetrische Verteilung des Dienstalters schliessen lässt.

Klassengrösse: Es befinden sich im Durchschnitt 19.33 Lernende pro Klasse mit einer Standardabweichung von 4.06 (unter Ausschluss der vier Ausreisser).

VA-Quote: Nach Ausschluss der vier Ausreisser sind gemäss den Angaben der Lehrpersonen durchschnittlich 10.48 Prozent der Schüler/-innen pro Klasse verhaltensauffällig.

Im Strukturgleichungsmodell geschätzte Abhängigkeitsbeziehungen zwischen dem Kriterium ‚Integrationsoptimismus‘ und vier Prädiktoren

Das im Folgenden dargestellte Strukturgleichungsmodell (vgl. Abbildung 1) wurde in AMOS 20 aufgesetzt und berechnet. Als Datengrundlage dienten zwei separate jedoch strukturidentische Rohdatensätze für die Klassenlehrpersonen der Primar- bzw. der Sekundarstufe. Sämtliche manifesten Variablen wurden z-transformiert, um möglichen Verzerrungen aufgrund der unterschiedlichen Skalierung vorzubeugen. Das Modell wurde auf der Grundlage des FIML-Verfahrens (Full Information Maximum Likelihood) geschätzt. Das heisst, fehlende Werte wurden von der Software anhand sämtlicher im Modell verfügbaren Informationen berechnet und hernach für die Parameterschätzung verwendet. Das Modell erreicht gute Fit-Werte ($\chi^2=73.42$, $df=69$, n.s.; RMSEA=0.02, $pclose=0.96$; TLI=0.98; CFI=0.99). Im Sample der Primarlehrpersonen erklären die vier Prädiktoren gemeinsam 26 Prozent der Varianz von ‚Integrationsoptimismus‘, im Sample der Sekundarlehrpersonen sind es 36 Prozent.

Das Modell wird in Abbildung 1 nach gängiger Konvention dargestellt: Die eckigen Kästen repräsentieren manifeste Variablen, Ovale stehen für latente Konstrukte, die anhand des Schätzverfahrens ermittelt wurden. Die Residuen der Schätzung sind als Kreissymbole

dargestellt (e1 – e7 bei den manifesten Parcels; d1 – d3 bei den latenten Konstrukten). Die unidirektionalen Pfeile repräsentieren gerichtete Einflüsse, die bidirektionalen (abgerundeten) Pfeile stehen dagegen für Korrelationen ohne theoretisch begründete Kausalrichtung. Direkt neben den Pfeilen sind die standardisierten Parameter (Faktorladungen, Regressions- und Korrelationskoeffizienten) abgebildet – der obere Wert gilt immer für die Primarlehrpersonen, der untere Wert für die Sekundarlehrpersonen.

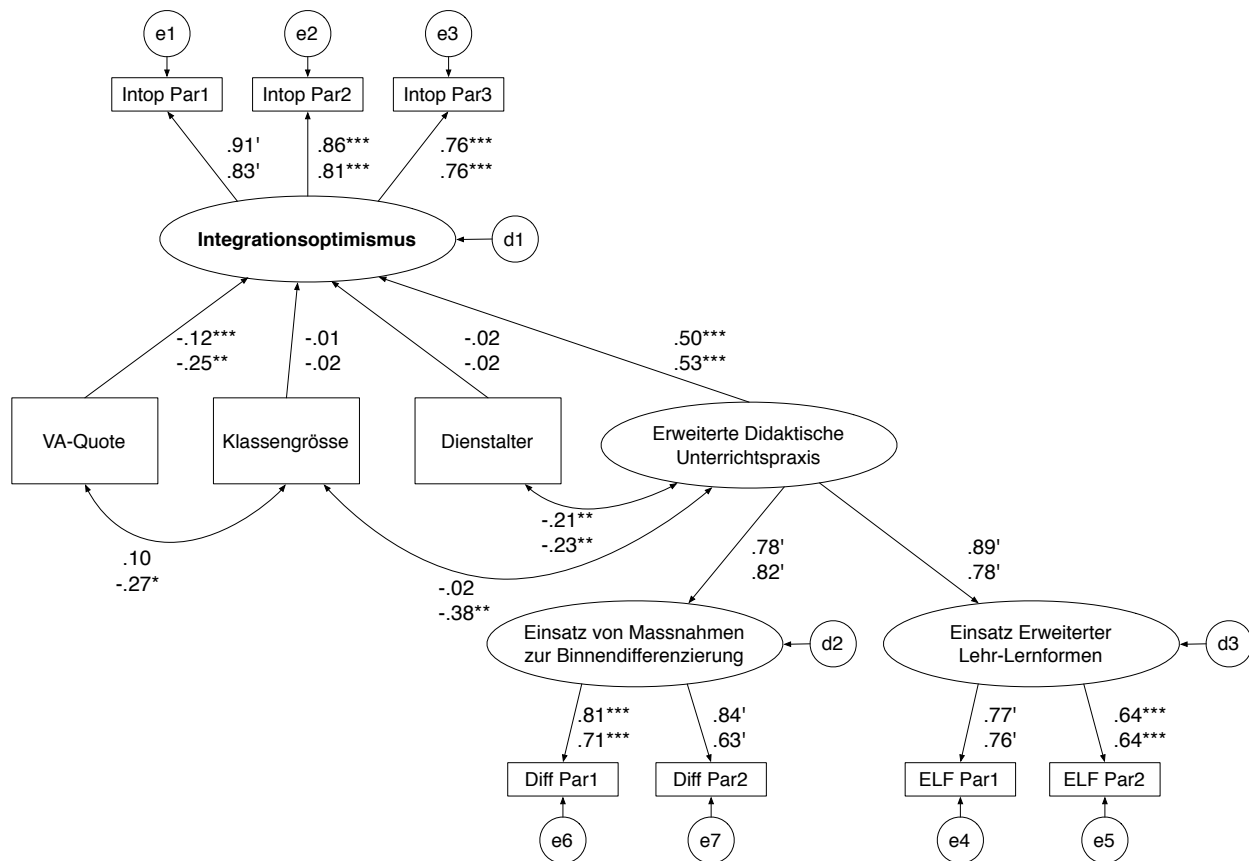


Abbildung 1: Prädiktion von 'Integrationsoptimismus' im Strukturgleichungsmodell

Legende der symbolisch dargestellten Signifikanzwerte in Abbildung 1

' Prüfung auf Signifikanz ist nicht möglich, da die unstandardisierte Faktorladung auf 1 fixiert wurde

* $p < 0.1$ Parameter ist auf dem 10 % Alphafehlerniveau signifikant

** $p < 0.05$ Parameter ist auf dem 5 % Alphafehlerniveau signifikant

*** $p < 0.001$ Parameter ist auf dem 0.1 % Alphafehlerniveau signifikant

Zunächst ist Folgendes festzuhalten: Die Modellierung impliziert, dass die tatsächliche Zuversicht der Lehrpersonen bezüglich der Umsetzung der integrativen Volksschule – im Sinne einer latenten Überzeugung – ihr manifestes Antwortverhalten während der Onlinebefragung vorhersagte. Die Werte der drei Parcels von ‚Integrationsoptimismus‘ (Intop Par1, Intop Par2, Intop Par3) fungieren deshalb bei der Berechnung des latenten Konstrukts als abhängige Variablen. Diese Kausalrichtung ist in Abbildung 1 anhand der drei Pfeile dargestellt, die unidirektional zu den drei Parcels hinführen. Das Schätzverfahren beruht auf dem Prinzip der linearen Regression, wobei die Beta-Gewichte in diesem Fall als Faktorladungen bezeichnet werden. Deren Wertebereich von 0.76 bis 0.91 deutet an, dass das latente Konstrukt das manifeste Antwortverhalten gut vorhersagt. Der Vorteil dieses Vorgehens besteht darin, dass ‚Integrationsoptimismus‘ als latenter Factorscore messfehlerkorrigiert geschätzt wird, indem die Residuen der Schätzung (e_1 , e_2 , e_3) explizit berücksichtigt werden.

Der messfehlerbereinigte, latente Factorscore ‚Integrationsoptimismus‘ dient dem Modell als Kriteriumsvariable zur Schätzung des Einflusses der vier Prädiktoren. Diese Kausalrichtung wird in Abbildung 1 anhand der vier unidirektionalen Pfeile dargestellt, die zum Kriterium ‚Integrationsoptimismus‘ hinführen. Die standardisierten Regressionskoeffizienten können anhand der acht abgebildeten Werte stufenspezifisch abgelesen werden.

Der mit Abstand *einflussreichste Prädiktor ist die ‚Erweiterte Didaktische Unterrichtspraxis‘* (EDUP), die als latentes Konstrukt zweiter Ordnung modelliert wurde. Mit dieser Modellierung wird der gemeinsame Anteil an Varianz der messfehlerbereinigten latenten Konstrukte erster Ordnung im übergeordneten Konstrukt EDUP gebündelt⁴. Massnahmen zur Binnendifferenzierung und der Einsatz Erweiterter Lehr-Lernformen tragen dank diesem

Vorgehen also etwa gleichviel zur Erklärung von Integrationsoptimismus bei – zusammen sind dies 25 Prozent (Primarlehrpersonen) bzw. 28 Prozent (Sekundarlehrpersonen).

Alle anderen Prädiktoren – ‚VA-Quote‘, ‚Klassengrösse‘, ‚Dienstalter‘ – wurden als manifeste Variablen ohne Messfehlerkorrektur im Modell integriert. Die Quote verhaltensauffälliger Schüler/-innen hat sich dabei als zweitstärkster Prädiktor erwiesen, der für sich alleine knapp 1.5 Prozent (Primarlehrpersonen) bzw. etwas mehr als 6 Prozent (Sekundarlehrpersonen) Varianz von Integrationsoptimismus aufklärt. Von der Klassengrösse und vom Dienstalter gehen indes keine bedeutsamen Effekte aus.

Aufgrund der theoretischen Annahme, dass zwischen diesen Variablen vielfältige, wechselwirksame Beziehungen vorherrschen, wurde zunächst zwischen alle Prädiktoren Korrelationen geschätzt. Dabei hat sich herausgestellt, dass die Quote verhaltensauffälliger Schüler/-innen weder mit dem Dienstalter der Lehrpersonen noch mit ihrer didaktischen Unterrichtspraxis in einem statistisch signifikanten Zusammenhang steht (diese nicht signifikanten Korrelationen wurden für die Schätzung des berichteten Modells nicht mehr berücksichtigt, d.h. es erfolgte in allen entsprechenden Fällen und für beide Teilgruppen eine Restriktion auf eine Nullkorrelation).

Alle berichteten Parameter wurden anhand eines Chi²-Differenzentests auf Messinvarianz zwischen den zwei Gruppen der Primar- bzw. Sekundarlehrpersonen überprüft. Aufgrund der strukturell unterschiedlichen Rahmenbedingungen dieser zwei Gruppen wurde vermutet, einige statistisch signifikante Unterschiede zu identifizieren. Das war aber kaum der Fall. Die einzigen bedeutsamen Unterschiede fanden sich in den Korrelationen zwischen ‚VA-Quote‘ und ‚Klassengrösse‘ sowie zwischen ‚Klassengrösse‘ und EDUP. Das Modell wurde entsprechend restringiert: Es wurden alle Parameter, bei denen der Chi²-Differenzentests nicht signifikant

ausfiel, für die beiden Gruppen als gleich hoch definiert. Dies ist in Abbildung 1 nicht nachvollziehbar, weil die standardisierten Werte abgedruckt sind, die sich zwischen den Gruppen unterscheiden können, obwohl die unstandardisierten gleichgesetzt worden sind.

Diskussion

Reflexion der Ergebnisse des Strukturgleichungsmodells

Im berichteten Strukturgleichungsmodell wurde der Einfluss ausgewählter Prädiktoren auf das Kriterium ‚Integrationsoptimismus‘ geschätzt. Der wichtigste Befund dieser Kausalanalyse betont die *Bedeutsamkeit einer erweiterten didaktischen Unterrichtspraxis* als Faktor einer positiven Einstellung gegenüber der Implementation der integrativen Schulform. Lehrpersonen, die regelmässig und routiniert binnendifferenziert und auf der Grundlage Erweiterter Lehr-Lernformen unterrichten, äussern vergleichsweise weniger Hürden und Stolpersteine im Umgang mit Heterogenität. Im Unterschied zu Lehrpersonen, die in geringerem Umfang binnendifferenzierend und nach erweiterten Lernformen unterrichten sind diese zuversichtlicher, dass eine integrative Volksschule gewinnbringend realisiert werden kann. Dieser erste Hauptbefund findet im Strukturgleichungsmodell in einem entsprechend hohen standardisierten Regressionskoeffizienten, der für sich allein 25 Prozent bzw. 28 Prozent der Varianz von ‚Integrationsoptimismus‘ aufklärt, seinen Niederschlag.

Der zweitstärkste Prädiktor ist die Quote verhaltensauffälliger Schüler/-innen pro Klasse, welche den Integrationsoptimismus der Lehrpersonen erwartungsgemäss ungünstig beeinflusst. An dieser Variable ist bemerkenswert, dass sie nur mit der Klassengrösse signifikant korreliert. Die Tatsache, dass weder ein Zusammenhang zur erweiterten didaktischen Unterrichtspraxis noch zum Dienstalter der Lehrpersonen besteht, ist erklärungsbedürftig. Theoretisch wäre es

nicht unplausibel, dass dienstältere Lehrpersonen aufgrund ihres grossen Erfahrungsschatzes ein vergleichsweise effizientes Classroom-Management beherrschen. Folglich bestünde die Chance für Lernende mit Disposition zu Verhaltensauffälligkeit, dass sie im Rahmen eines solch effizient geführten Unterrichts vermehrt zur produktiven Mitarbeit angeregt werden. Ähnlich sollte es sich theoretisch mit der erweiterten didaktischen Unterrichtspraxis verhalten, weil sich ein solcher Unterricht gemäss der referierten Forschung bei Verhaltensauffälligkeit stabilisierend auswirkt. Das heisst, dass in Klassen von Lehrpersonen mit hoch ausgeprägter erweiterter didaktischer Unterrichtspraxis vergleichsweise weniger Kinder verhaltensauffällig sein sollten. Die fehlenden Korrelationen könnten somit eine gewisse Renitenz von Verhaltensauffälligkeit gegenüber den vermuteten pädagogisch-didaktischen Interventionen andeuten.

Oder es besteht ein Problem mit den verwendeten Daten aufgrund einer Formulierung im Erhebungsinstrument (Onlinefragebogen). Der Erhebungsplan sah vor, dass die Lehrpersonen aufgrund der explizit formulierten Frage 18 nach der Anzahl förderbedürftiger Kinder in der Klasse auch bei Frage 19 nach der Anzahl Verhaltensauffälligen nur diejenigen Lernenden dazu zählen, die einen offiziellen Förderstatus aufweisen. Vermutlich ist ein Grossteil der Lehrpersonen dieser Logik auch gefolgt. Aufgrund der ungenauen Formulierung ist aber nicht auszuschliessen, dass einige Lehrpersonen bei der Frage nach den Verhaltensauffälligen ihre persönliche Einschätzung zum Ausdruck brachten, also höhere Werte angaben, als tatsächlich Kinder ihrer Klasse einen offiziellen Förderstatus haben. Falls dies passiert ist, ist des Weiteren zu vermuten, dass es vor allem diejenigen Lehrkräfte anders als vorgesehen machten, für welche die Verhaltensauffälligen ohnehin ein Reizthema darstellen. Aufgrund dieser mutmasslich systematischen Verzerrung wäre der ermittelte, negative Effekt der Quote an Verhaltensauffälligen pro Klasse auf den Integrationsoptimismus der Lehrpersonen überschätzt,

weil die vermeintlich objektive VA-Quote mit der Gereiztheit bestimmter Lehrpersonen konfundiert ist.

Dass die Effekte des Dienstalters auf den Integrationsoptimismus nicht signifikant ausfallen, ist eher unerwartet, jedoch erklärbar: Das Dienstalter hängt systematisch mit der erweiterten Didaktischen Unterrichtspraxis zusammen ($r = -0.21$ bzw. -0.23). Demnach gestalten dienstältere Lehrpersonen ihren Unterricht tendenziell weniger binnendifferenziert bzw. auf der Grundlage der ELF als ihre jüngeren Kolleginnen und Kollegen. In Konkurrenz zur Unterrichtspraxis erweist sich somit das Dienstalter als relativ unbedeutender Prädiktor zur Erklärung von Integrationsoptimismus. Allerdings sind die negativen Korrelationen zwischen Dienstalter und Unterrichtspraxis als moderat anzusehen. Dies bedeutet, dass ein wesentlicher Teil der dienstälteren Lehrpersonen entgegen der Tendenz binnendifferenziert und nach den Prinzipien erweiterter Lehr-Lernformen unterrichten. Und umgekehrt: Nicht alle jungen Lehrpersonen pflegen eine ausgeprägte erweiterte didaktische Unterrichtspraxis.

Der nicht signifikante Einfluss der Klassengrösse auf die Zuversicht der Lehrpersonen reiht sich nahtlos in die uneinheitliche Befundlage des referierten Forschungsstandes ein. Immerhin deutet das negative Vorzeichen an, dass von grossen Klassen kein Anstieg des Integrationsoptimismus zu erwarten ist. Interessant sind vor allem die negativen Korrelationen mit der erweiterten didaktischen Unterrichtspraxis bzw. mit der Quote Verhaltensauffälliger, die nur auf der Sekundarstufe signifikant ausfallen. Der negative Zusammenhang zwischen Klassengrösse und VA-Quote ($r = -0.27$) bedeutet, dass in den kleineren Sekundarklassen verhältnismässig viele Schüler/-innen verhaltensauffällig sind. Insofern kann spekuliert werden, dass die Verhaltensauffälligen mehrheitlich in den kleinen Sek-B resp. in den noch kleineren Sek-C Klassen beschult werden⁵. Des Weiteren deutet die negative Korrelation mit der

erweiterten didaktischen Unterrichtspraxis ($r = -0.38$) an, dass Lehrpersonen kleinerer Sekundarklassen ihren Unterricht eher binnendifferenziert und auf der Grundlage der ELF gestalten. Die Unterrichtsgestaltung erweist sich also erneut als der bedeutsamere Prädiktor, sodass die Klassengrösse keinen signifikanten Beitrag zur Varianzaufklärung von Integrationsoptimismus leistet.

Möglichkeiten und Grenzen der SEM-Analysen

Das Strukturgleichungsmodell erreicht gute Fit-Werte und die geschätzten Parameter entsprechen mehrheitlich den theoretischen Erwartungen, was insgesamt für die Stabilität des Modells spricht. Allerdings werden rund zwei Drittel der Varianz von ‚Integrationsoptimismus‘ durch die vier modellierten Prädiktoren nicht erklärt. Dieser Anteil unerklärter Varianz ist im Vergleich zu anderen sozialwissenschaftlichen Studien nicht besonders gross. Dennoch wäre es im Sinne aller Beteiligten wünschbar, wenn die erziehungswissenschaftliche Forschung noch besser erklären könnte, wie die Überzeugungen von Lehrpersonen bezüglich der integrativen Volksschule zustande kommen. Insbesondere sind weitere Bedingungen denkbar, die im Rahmen der berichteten Analysen nicht berücksichtigt wurden (z.B. strukturelle Rahmenbedingungen auf der Ebene der Einzelschulen, personenbezogene Merkmale der Lehrpersonen).

Des Weiteren sind die berichteten *Kausalrichtungen* zu problematisieren (z.B. Unterrichtspraxis prädiziert Integrationsoptimismus), die anhand der verfügbaren Querschnittsdaten empirisch nicht abgesichert werden können. Theoretisch wären auch umgekehrte Kausalrichtungen denkbar, was empirisch aber nur mit Längsschnittdaten und allenfalls mit experimentellen Designs zu überprüfen wäre.

Schliesslich ist als methodologischer Vorbehalt nochmals zu erwähnen, dass die Stichprobenziehung nicht nach dem Zufallsprinzip erfolgte. Folglich sind Rückschlüsse auf die Grundgesamtheit der Zürcher Lehrpersonen nicht zulässig. Für die Analysen des vorliegenden Artikels schien es dennoch angebracht, inferenzstatistische Verfahren anzuwenden, um anhand der Signifikanzwerte der geschätzten Parameter die Plausibilität dieser Schätzungen zu überprüfen.

Resümee

Ideelle Wertschätzung vs. pragmatische Skepsis: Ein Spannungsfeld zeichnet sich ab

Obwohl die Lehrpersonen mehrheitlich positiv überzeugt sind von dem bildungspolitischen Konzept der integrativen Schulform und beispielsweise verschiedene Vorteile für die integrierten Kinder mit besonderen pädagogischen Bedürfnissen anerkennen (handlungsferne Überzeugungen), äussern sie sich in der Gesamtbilanz tendenziell skeptisch, wenn sie auf die konkrete Umsetzung der integrativen Volksschule angesprochen werden (handlungsnahe Überzeugungen). Denn viele Lehrkräfte finden, dass ihre Arbeit im Vergleich zum ehemals separativen Schulsystem mit Mehraufwand verbunden und anspruchsvoller geworden sei – ohne gesteigerte Erfolgsaussichten. Im vorliegenden Artikel wurde herausgearbeitet, dass unterschiedliche Voraussetzungen dazu beitragen, dass gewisse Lehrpersonen optimistischer sind als andere betreffend die Realisierbarkeit der integrativen Volksschule. Beispielsweise sind Lehrkräfte der Primarstufe zuversichtlicher als Sekundarlehrpersonen.

Binnendifferenzierung und Erweiterte Lehr-Lernformen als wichtige Gelingensbedingungen

Erfolgsversprechende Massnahmen für den Umgang mit heterogenen Lerngruppen sind die Binnendifferenzierung des Unterrichts sowie der Einsatz Erweiterter Lehr-Lernformen. Eine Unterrichtsgestaltung, welche diesen Prinzipien folgt, gehört zu den wichtigsten Bedingungen der Zuversicht von Lehrpersonen betreffend die Umsetzung der integrativen Volksschule. Folglich ist zu überlegen, mit welchen Angeboten der Aus- und Weiterbildung entsprechende didaktische Kompetenzen nachhaltig vermitteln werden können, damit den Lehrpersonen der Umgang mit Heterogenität leichter fällt.

Verhaltensauffälligkeit und strukturelle Rahmenbedingungen als wichtige Einflussgrössen

Der zweitwichtigste Befund der SEM-Analysen, wonach wahrgenommene Verhaltensauffälligkeit die Zuversicht der Lehrer/-innen einer gelingenden integrativen Volksschule in bedeutsamer Weise schmälert, ist erwartungskonform. Erklärungsbedürftig bleibt jedoch, unter welchen Voraussetzungen die Lehrpersonen vergleichbare Handlungen der Schüler/-innen *als auffällig oder störend wahrnehmen*. Entsprechend sind weiterführende Forschungsprojekte zum Themenkreis ‚Verhaltensauffälligkeit/Unterrichtsstörungen‘ erforderlich.

Zu resümieren ist des Weiteren, dass in Konkurrenz zur Unterrichtspraxis kein nennenswerter Einfluss der Klassengrösse und des Dienstalters der Lehrpersonen ermittelt wurde. Hierzu gilt es zu überprüfen, ob mit anderen Daten und/ oder Methoden vielleicht doch kritisch Grenzwerte identifiziert werden können. Auf jeden Fall braucht es mehr Forschung, um

den Einfluss der strukturellen Rahmenbedingungen (z.B. auf der Ebene der Einzelschulen) in Interaktion mit der Unterrichtspraxis auf die Genese der Überzeugungen von Lehrpersonen bezüglich Integration besser zu verstehen.

Anmerkungen

- 1 Die grundsätzliche Idee des Parcelings besteht darin, SEM-Analysen nicht mit zu vielen zu schätzenden Informationen zu belasten. Das Zusammenfassen der 6 Items zu 3 Parcels von ‚Integrationsoptimismus‘ bedeutet eine Halbierung der zu schätzenden Parameter.
- 2 Sämtliche explorativen Faktorenanalysen wurden mit SPSS 20 berechnet. Die Werte des KMO-Kriteriums lagen jeweils über 0.8. Als Extraktionsmethode wurde immer die Hauptkomponentenanalyse mit obliquer Rotationsmethode gewählt. Als Kriterium zur Bestimmung der Faktorenzahl diente das Kaiser-Guttman-Kriterium (Eigenwerte > 1.0).
- 3 Zur Berechnung der VA-Quote wurde die absolute Anzahl Verhaltensauffälliger geteilt durch die Klassengrösse. Dadurch wurde ein Wertebereich von 0-1 definiert, der multipliziert mit dem Faktor 100 den prozentualen Anteil Verhaltensauffälliger pro Schulklasse ausweist
- 4 Wie im Zuge konfirmatorischer Faktorenanalysen bei der Vorbereitung des berichteten Strukturgleichungsmodells ermittelt wurde, sind die zwei latenten Konstrukte erster Ordnung hochgradig miteinander korreliert ($r > 0.6$). Das bedeutet, dass Lehrpersonen, die häufig ELF einsetzen, den Unterricht häufig auch binnendifferenziert gestalten. Würden diese zwei Komponenten als eigenständige Prädiktoren im Modell eingeführt, stünden sie aufgrund der hohen Korrelation bei der Varianzaufklärung von ‚Integrationsoptimismus‘ in starker Konkurrenz zueinander und ihr gemeinsamer Anteil an erklärter Varianz würde nicht sichtbar. Deshalb wurde die Modellierung des latenten Konstrukts 2. Ordnung präferiert. Um den Überschneidungsbereich auf beide Komponenten etwa gleich abzustützen, wurden die beiden unstandardisierten Faktorladungen von EDUP auf die zwei Konstrukte 1. Ordnung auf den Wert 1 fixiert.
- 5 Die Vermutung basiert auf den offiziellen Angaben der Bildungsstatistik zur Klassengrösse der verschiedenen Abteilungen der Sekundarstufe (<http://www.bista.zh.ch/vs/SHaus.aspx> [19.07.2012]). Es wurden keine eigenen Berechnungen vorgenommen.

Literatur

- Altrichter, H., Trautmann, M., Wischer, B., Sommerau, S. & Doppler, B. (2009). Unterrichten in heterogenen Gruppen. Das Qualitätspotenzial von Individualisierung, Differenzierung und Klassenschülerzahl. In W. Specht (Hrsg.), *Nationaler Bildungsbericht Österreich 2009* (2 Bde). Graz: Leykam.
- Bentler, P. M. (2006). *EQS 6 Structural equations program manual*. Encino, CA: Multivariate Software.
- Blatchford, P., Bassett, P. & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher-pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction*, 21, 715-730.
- Bohl, T., Batzel, A. & Richey, P. (2011). Öffnung - Differenzierung - Individualisierung - Adaptivität. Charakteristika, didaktische Implikationen und Forschungsbefunde verwandter Unterrichtskonzepten zum Umgang mit Heterogenität. *Schulpädagogik heute*, 2 (4), 40-71.
- Bohl, T. & Kucharz, D. (2010). *Offener Unterricht heute. Konzeptionelle und didaktische Weiterentwicklung*. Weinheim: Beltz.
- Brühwiler, C. & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21, 95-108.
- Buholzer, A. & Kummer Wyss, A. (2010). *Alle gleich - alle unterschiedlich! Zum Umgang mit Heterogenität in Schule und Unterricht*. Seelze: Kallmeyer.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS basic concepts, applications and programming* (2nd ed.). New York, NY: Routledge.
- Croci, A., Imgrüth, P., Landwehr, N. & Spring, K. (1995). *ELF: Ein Projekt macht Schule: Magazin zum Thema erweiterte Lernformen*. Littau, Buchs: Kantonaler Lehrmittelverlag Luzern, Kantonaler Lehrmittelverlag Aargau.
- Curran, P. J., West, S. G. & Finch, J. F. (1996). The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis. In R. H. Hoyle (Hrsg.), *Structural Equation Modelling. Concepts, Issues, and Applications* (S. 16-29). Thousand Oaks, CA: Sage Publications.
- Gebhardt, M., Schwab, S., Reicher, H., Ellmeier, B., Gmeiner, S., Rossmann, P. et al. (2011). Einstellungen von LehrerInnen zur schulischen Integration von Kindern mit einem sonderpädagogischen Förderbedarf in Österreich. *Empirische Sonderpädagogik* (4), 275-290.
- Gröhlich, C., Scharenberg, K. & Bos, W. (2009). Wirkt sich Leistungsheterogenität in Schulklassen auf den individuellen Lernerfolg in der Sekundarstufe aus? *Journal for Educational Research Online / Journal für Bildungsforschung Online*, 1 (1), 86–105.
- Harfitt, G. J. (2012). An examination of teachers' perceptions and practice when teaching large and reduced-size classes: Do teachers really teach them in the same way? *Teaching and Teacher Education*, 28 (1), 132–140.
- Hartinger, A. (2005). Verschiedene Formen der Öffnung von Unterricht und ihre Auswirkung auf das Selbstbestimmungsempfinden von Grundschulkindern. *Zeitschrift für Pädagogik*, 51 (3), 397-414.
- Joller-Graf, K. (2010). Binnendifferenziert unterrichten. In A. Buholzer & A. Kummer-Wyss (Hrsg.), *Alle gleich - alle unterschiedlich! Zum Umgang mit Heterogenität in Schule und Unterricht*. Seelze-Velber: Kallmeyer.

- Kantonsrat des Kantons Zürich. (2005). Volksschulgesetz (VSG).
- Klieme, E. & Warwas, J. (2011). Konzepte der individuellen Förderung. *Zeitschrift für Pädagogik*, 57 (6), 805-818.
- Kunz, A., Luder, R. & Moretti, M. (2010). Die Messung von Einstellungen zur Integration (EZI). *Empirische Sonderpädagogik* (3), 83-94.
- Liesen, C. & Luder, R. (2011). Forschungsstand zur integrativen und separativen schulischen Förderung von Schülerinnen und Schülern mit Verhaltensauffälligkeiten. *Schweizerische Zeitschrift für Heilpädagogik*, 17 (8), 11-18.
- Lipowsky, F. (2002). Zur Qualität offener Lernsituationen im Spiegel empirischer Forschung - auf die Mikroebene kommt es an. In U. Drews & W. Wallrabenstein (Hrsg.), *Freiarbeit in der Grundschule. Offener Unterricht in Theorie, Forschung und Praxis* (S. 126-159). Frankfurt, Main: Grundschulverband.
- Pauli, C., Reusser, K. & Grob, U. (2010). Reformorientierter Mathematikunterricht in der Deutschschweiz. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität* (S. 309-340). Münster: Waxmann.
- Pauli, C., Reusser, K., Waldis, M. & Grob, U. (2003). Erweiterte Lehr- und Lernformen im Mathematikunterricht der Deutschschweiz. *Unterrichtswissenschaft*, 31 (4), 291-320.
- Prenzel, A. (2006). *Pädagogik der Vielfalt: Verschiedenheit und Gleichberechtigung in Interkultureller, Feministischer und Integrativer Pädagogik*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Preuss-Lausitz, U. & Textor, A. (2006). Das «Emsoz-Projekt»: Zur Integration schwieriger Kinder in der Grundschule. In *Inklusive Bildungsprozesse*. (S. 132-139). Bad Heilbrunn: Klinkhardt.
- Regierungsrat des Kantons Zürich. (2006). Volksschulverordnung (VSV) des Kantons Zürich vom 28. Juni 2006 (Bd. 2012).
- Reusser, K. & Stebler, R. (2000). Progressive, classical or balanced - a look at mathematical learning environments in Swiss-German lower-secondary schools. *Zentralblatt für Didaktik der Mathematik*, 32 (1), 1-10.
- Reusser, K. (2008). Empirisch fundierte Didaktik - didaktisch fundierte Unterrichtsforschung. Eine Perspektive zur Neuorientierung der Allgemeinen Didaktik. In *Perspektiven der Didaktik*. (S. 219-237). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Reusser, K. (2011). Unterricht und Klassenführung. In L. Criblez, B. Müller & J. Oelkers (Hrsg.), *Die Volksschule zwischen Innovationsdruck und Reformkritik* (S. 68-83). Zürich: Neue Zürcher Zeitung.
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität - Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität* (S. 9-32). Münster: Waxmann.
- Reusser, K., Pauli, C. & Elmer, A. (2011). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 478-495). Münster: Waxmann.
- Reusser, K., Stebler, R., Mandel, D. & Eckstein, B. (2013). *Erfolgreicher Unterricht in heterogenen Lerngruppen auf der Volksschulstufe des Kantons Zürich. Wissenschaftlicher Bericht zu Handen der Bildungsdirektion des Kantons Zürich*.
- Schukajlow, S. & Blum, W. (2011). Zum Einfluss der Klassengrösse auf Modellierungskompetenz, Selbst- und Unterrichtswahrnehmungen von Schülern in

- selbstständigkeitsorientierten Lehr-Lernformen. *Journal für Mathematik-Didaktik*, 32, 133-151.
- Seidel, T. & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade. The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of educational research*, 77 (4), 454-499.
- Tillmann, K.-J. & Wischer, B. (2006). Heterogenität in der Schule. Forschungsstand und Konsequenzen. *Pädagogik*, 58 (3), 44-48.
- Trautmann, M. & Wischer, B. (2008). Das Konzept der Inneren Differenzierung - eine vergleichende Analyse der Diskussion der 1970er Jahre mit dem aktuellen Heterogenitätsdiskurs. In M. A. Meyer, M. Prenzel & S. Hellekamps (Hrsg.), *Perspektiven der Didaktik*. (S. 159-172). Wiesbaden: VS Verlag für Sozialwissenschaft.
- Trautmann, M. & Wischer, B. (2011). *Heterogenität in der Schule. Eine kritische Einführung*. 1. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Voss, T., Kleickmann, T., Kunter, M. & Hachfeld, A. (2011). Überzeugungen von Mathematiklehrkräften. In *Professionelle Kompetenz von Lehrkräften*. (S. 235-257). Münster u.a.: Waxmann.

6.1.2. Unterrichtliche Devianz und subjektives Störungsempfinden³¹

Abstract

Unterrichtsstörungen gelten als eine der belastendsten Herausforderungen in der schulpädagogischen Praxis, was auch in der gegenwärtigen Inklusionsdebatte mit teilweise neuen Akzenten thematisiert wird. Dennoch fehlt es an profundem Wissen zum subjektiven Störungsempfinden von Lehrpersonen Schülerinnen und Schülern³² sowie zur Bedeutung von Störungskontexten. Diese Forschungslücke aufgreifend wurde in der Studie zur Untersuchung gestörten Unterrichts (SUGUS) auf der Basis eines interaktionistischen Theorierahmens ein mehrperspektivisch angelegtes Instrumentarium zur Erfassung von Unterrichtsstörungen entwickelt und einem Pretest mit 11 Klassen des 5. Schuljahrs unterzogen. Präsentiert werden die faktorielle Struktur der Instrumente mittels eines Mehrebenen-Strukturgleichungsmodells und deskriptive Kennwerte zu den theoretischen Konstrukten.

1 Problemaufriss

In der (fach-)öffentlichen Inklusionsdebatte gelten Kinder und Jugendliche mit besonderen pädagogischen Bedürfnissen im emotional-sozialen Bereich als die herausforderndste Schülergruppe (Liesen & Luder, 2011). Die Forschungslage deutet darauf hin, dass v.a. eine externalisierende Problemverarbeitung im Unterricht des Regelschulbetriebs von

³¹ Copyright © 2016 by Verlag Empirische Pädagogik. Reproduced [or adapted] with permission. The official citation that should be used in referencing this material is: Eckstein, B., Grob, U. & Reusser, K. (2016). Unterrichtliche Devianz und subjektives Störungsempfinden. Entwicklung eines Instrumentariums zur Erfassung von Unterrichtsstörungen. *Empirische Pädagogik*, 30(1), 113-129.

³² In der Originalarbeit wurde gemäss den Vorgaben des Zeitschriftenverlags das generische Maskulinum verwendet ("Schüler" für Jungen *und Mädchen*). Um der Gender Policy der Universität Zürich zu entsprechen wird dies in der vorliegenden Fassung angepasst.

den meisten Lehrpersonen sowie von manchen Mitschülerinnen und Mitschülern als störend empfunden wird.

Allerdings bestehen Unterrichtsstörungen nicht per se, sondern werden immer subjektiv im Kontext eines didaktischen Settings wahrgenommen, gedeutet und bewertet (Wettstein & Thommen, 2007). Wenn bspw. ein Schüler oder eine Schülerin den Platz verlässt oder Zwischengespräche führt, weist dies im Klassenunterricht aufgrund dessen rigiden Verhaltensanforderungen und Kommunikationsnormen (Zevenbergen, 2001) ein größeres Störungspotenzial auf als während einer Gruppenarbeit; sodann empfinden nicht alle Lehrpersonen solche Situationen als gleichermaßen störend.

Ausgehend von diesem Problemverständnis werden in der *Studie zur Untersuchung gestörten Unterrichts* (SUGUS) zwei miteinander interagierende, kontextabhängige Merkmalsdimensionen von Unterrichtsstörungen unterschieden: Normabweichendes Schülerverhalten und subjektives Störungsempfinden. Ein langfristiges Ziel der SUGUS-Studie besteht darin, pädagogisch-didaktische Bedingungen zu identifizieren, unter denen normabweichendes Verhalten vergleichsweise selten auftritt und als wenig störend empfunden wird.

2 *Zentrale Merkmalsdimensionen von Unterrichtsstörungen*

2.1 *Unterrichtliche Devianz*

Alle Schüler/innen zeigen gelegentlich Verhaltensweisen, die von den situativ geltenden, teilweise latenten Normen im Unterricht abweichen (Mutzeck, 2000). Diejenigen Formen unterrichtlicher Devianz, welche am häufigsten vorkommen – z.T. mehrfach während einzelner Lektionen – stellen vergleichsweise bagatellhafte *Disziplinprobleme* dar, z.B.

„Schwatzen“ (Beaman, Wheldall & Kemp, 2007; Munn et al., 2013). *Aggressives Verhalten* gegenüber Mitschülerinnen oder Mitschülern sowie *oppositionelles Verhalten* gegenüber der Lehrperson kommen selbst auf der Oberstufe sehr selten vor (Müller, Fleischli & Hofmann, 2013).

In der Theorie und Forschung wird aus unterschiedlichen Perspektiven beschrieben, weshalb sich welche Kinder normabweichend verhalten (Myschker & Stein, 2014). Eine für die Schulpraxis und für die Ziele der SUGUS-Studie vielversprechende Suchlinie fokussiert auf die Aktualgenese unterrichtlicher Devianz, wobei „die prozessualen Merkmale des sozialen Miteinanders“ (Schönbächler, Herzog & Makarova, 2011, S. 322) bedeutsame Einflussgrößen darstellen, etwa die pädagogisch-didaktische Unterrichtsgestaltung (Hillenbrand, 2011): Während z.B. eine angemessene Strukturierung des Unterrichts in Kombination mit autonomieunterstützendem Lehrpersonenverhalten (Jang, Reeve & Deci, 2010; Textor, 2009) sowie binnendifferenzierenden Maßnahmen (Wettstein, Thommen & Eggert, 2010) mit geringer Häufigkeit unterrichtlicher Devianz einhergehen, verhalten sich viele Schüler/innen bei Über- oder Unterforderung vermehrt deviant (Schönbächler et al., 2009). Unterrichtliche Devianz konstituiert für sich alleine aber keine Unterrichtsstörung.

2.2 *Subjektives Störungsempfinden*

Eine Unterrichtsstörung liegt vor, wenn Lehrpersonen oder Mitschüler/innen deviantes Schülerverhalten mit Bezug auf ihre normativen Erwartungshorizonte als ablenkend, provokativ, belastend, usw. wahrnehmen, deuten und bewerten. Dieses subjektive Störungsempfinden interagiert mit personalen und kontextuellen Merkmalen und ist daher anfällig für *Wahrnehmungsverzerrungen* (Schweer & Thies, 2000). So weisen Lehrpersonen mit geringen

Selbstwirksamkeitsüberzeugungen eine erhöhte Störungssensitivität auf (Arbuckle & Little, 2004; Grosse Siestrup, 2010) und bagatellhafte Normabweichungen werden in „Wenigstörerklassen“ als störender empfunden als in „Vielstörerklassen“ (Makarova, Herzog & Schönbächler, 2014).

Weiter folgt das subjektive Störungsempfinden zeitlichen Verlaufsmustern. So etwa, wenn Lehrpersonen Verhaltensweisen von aufgrund früherer Erfahrungen mental als „Störer“ kategorisierten Schülerinnen oder Schülern verstärkt negativ wahrnehmen und sie in vergleichbaren Situationen progredient negativer beurteilen und behandeln als deren Mitschüler/innen (Hofer, 1986). Dies kann stigmatisierend wirken und ungünstige Rückkopplungseffekte nach sich ziehen, wie sie im Labeling Approach als *Aufschaukelungsprozess* beschrieben werden (Holtappels, 2000; Ulich, 1980). Solche Prozesse sind nachteilig für die betroffenen Kinder, sie belasten die Lehrpersonen und stellen für diese im Extremfall ein Gesundheitsrisiko dar (Krause, Dorsemagen & Alexander, 2011).

Schüler/innen empfinden unterrichtliche Devianz generell als weniger störend als die Lehrpersonen (Pfitzner & Schoppek, 2000). Viele können sich zwar aufgrund devianten Verhaltes ihrer Mitschüler/innen *weniger gut konzentrieren und manche ärgern sich* auch darüber, was mittelfristig die Qualität ihrer Beziehung zu den „Störern“ beeinträchtigen kann (Schönbächler et al., 2009). Aber einige Schüler/innen fühlen sich von unterrichtlicher Devianz überhaupt nicht gestört (ebd.) – manche finden solche Ereignisse sogar unterhaltsam (Breidenstein, 2006).

Die Forschungslage bilanzierend ist festzuhalten, dass Studien zum Wahrnehmungsgeschehen im Unterricht selten sind (Makarova, Herzog & Schönbächler, 2014) –

besonders mit spezifischem Blick auf pädagogisch-didaktische Bedingungen des Störungsempfindens.

2.3 Interaktionistisches Theoriemodell

Die gegenwärtige Theoriebildung integriert unterschiedliche Erklärungsansätze der Produktion und Rezeption von Unterrichtsstörungen: Distale und proximale Personen- und Kontextmerkmale bedingen sowohl die Aktualgenese unterrichtlicher Devianz als auch das darauf bezogene, subjektive Störungsempfinden von Lehrpersonen, Mitschülerinnen und Mitschülern (Stein & Stein, 2014). Abbildung 1 veranschaulicht dieses *interaktionistische Problemverständnis* in einem Theoriemodell, worin wechselwirksame Prozesse der interpersonalen Wahrnehmung und Beeinflussung in Anlehnung an Nickel (1985) dargestellt werden.

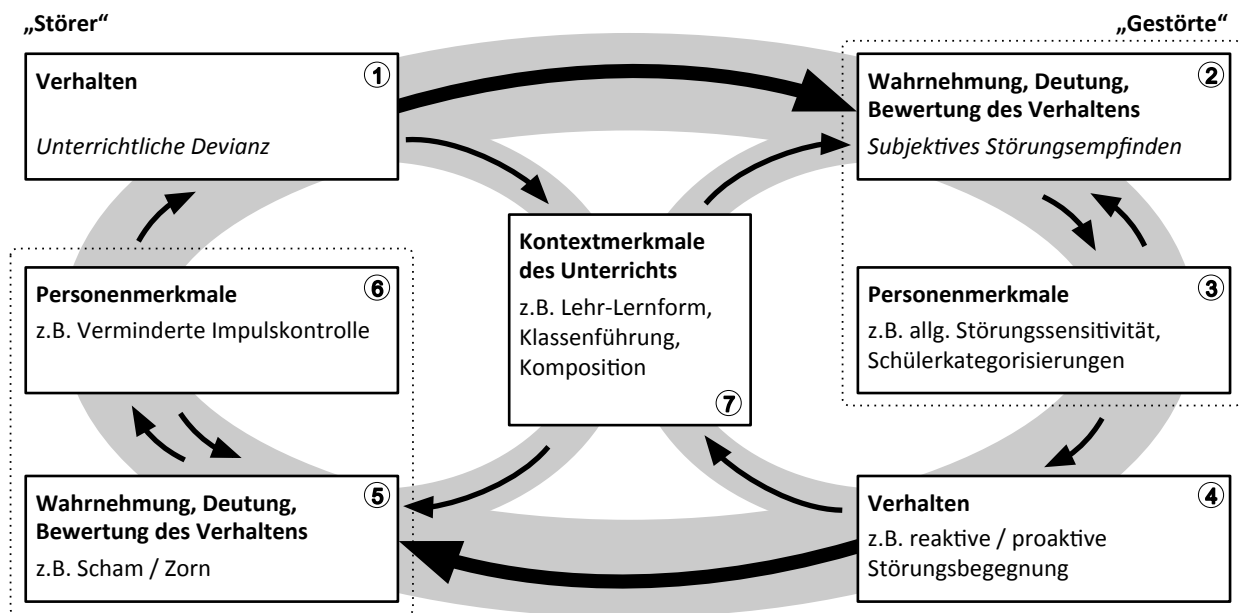


Abbildung 1: Produktion und Rezeption gestörten Unterrichts

Dieses Modell impliziert mannigfaltige Verlaufsmuster, wovon im Folgenden ein *beispielhafter Ablauf* skizziert sei: Ein Kind, das während einer Stillarbeit mit einem Lineal auf den Tisch trommelt, verstößt gegen das Ruhegebot dieser Unterrichtssequenz (1) – damit wird es zum potenziellen „Störer“. Bedingt wird dieses Verhalten u.a. von seiner Disposition, die ihm langes Stillsitzen erschwert (6). Die Lehrperson hat dieses Kind mental längst als „Zappelphilipp“ (Werning, 2008) eingeordnet (3) und beobachtet es im Unterricht sehr genau. Sie empfindet bereits den ersten Trommelschlag, den sie bei einem „guten Schüler“ oder bei einer „guten Schülerin“ vielleicht noch überhören würde, als störend (2). Damit wird die Lehrperson zum „Gestörten“ und die Situation als Ganzes zur Unterrichtsstörung.

Weil die Lehrperson akustische Unruhe im Unterricht generell als stark ablenkend erachtet (3), reagiert sie im Vergleich zu weniger lärmempfindlichen Kollegen früher mit einer pädagogischen Intervention: „Hör auf damit!“ (4). Diese Reaktion wird wiederum vom „Störer“ kognitiv und emotional gedeutet und verarbeitet (5). In der Folge sind sowohl kontraproduktive Aufschaukelungsprozesse als auch günstige Entwicklungen möglich, etwa wenn die Lehrperson mittelfristig gewisse Elemente des didaktischen Kontexts (7) anpasst und damit das situativ geltende Normensystem (z.B. das Ruhegebot während der Stillarbeit) verändert – mit zu vermutenden Wirkungen auf das Auftreten und die Bewertung von Störungen.

3 *Methodische Implikationen und Fragestellung*

Unterrichtsstörungen werden im (fach-)öffentlichen Diskurs häufig auf Probleme einzelner Schüler/innen zurückgeführt. Diese personalisierende Sichtweise berücksichtigt einerseits die aktuelle Theoriebildung nur ungenügend, andererseits gehen damit oft forschungsmethodische Probleme einher, etwa wenn ungenügend kontrollierte Rater- oder

Kontexteffekte die Einschätzungen der Informanten konfundieren (Korsch & Petermann, 2012). In der SUGUS-Studie werden verschiedene Maßnahmen ergriffen, um solche Effekte sichtbar zu machen bzw. nach Möglichkeit zu kontrollieren.

Zunächst wird unterrichtliche Devianz *niedrig-inferent*, anhand konkreter Verhaltensweisen mittels Häufigkeitsangaben für einen definierten Zeitraum erfasst, um auf Interpretationsspielräume des Instruments zurückführbare Ratereffekte zu minimieren. Zudem werden zielkindspezifische Einschätzungen eingeholt, da Ratings mit Bezug auf ganze Schulklassen eine gedankliche Aggregationsleistung verlangen, bei der verschiedene Informanten unterschiedliche Aspekte berücksichtigen können. Schließlich wird das *subjektive Störungsempfinden der Informanten explizit erfasst*, um es gemeinsam mit ihrer Einschätzung der Auftretenshäufigkeit unterrichtlicher Devianz zu modellieren. Ein langfristiges Ziel dieser Maßnahmen besteht darin, Bedingungen des subjektiven Störungsempfindens unter Kontrolle des gezeigten Verhaltens zu untersuchen.

Vorgesehen ist der Einsatz eines Fragebogens, welcher diese Kriterien erfüllt und zum Alter der Zielgruppe (5. Schuljahr) passt. Da kein solches Instrument gefunden wurde, ist ein neuer Fragebogen entwickelt und einem Pretest unterzogen worden, welcher folgender Fragestellung nachgeht:

Eignen sich die konstruierten Skalen zur Erfassung von unterrichtlicher Devianz und subjektivem Störungsempfinden zur Untersuchung von Unterrichtsstörungen?

4 Stichprobe, Datenerhebung und Analysestrategie

Im Sommer 2014 wurde im Kanton Zürich eine schriftliche Befragung mit 11 Klassen der 5. Primarschulstufe (11 Klassenlehrpersonen; 239 Schüler/innen, 10-11 Jahre) zu zwei

Zeitpunkten durchgeführt. Im Vorfeld der Erhebung bestimmte jede Lehrperson sechs Zielkinder aus ihrer Klasse anhand einer gezielten subjektiven Kontrastierung (zwei Kinder, die ihr mit Bezug auf das Verhalten positiv, und zwei, die ihr negativ auffallen, sowie zwei unauffällige, neutrale Kinder). Damit wurde ein hinreichendes Ausmaß an Merkmalsvarianz sichergestellt ohne übermäßigen Aufwand für den Pretest zu betreiben (kein vorgängiges Screening; reduzierte Anzahl einzuholender Ratings). So entstand eine nicht-repräsentative Teilstichprobe mit 66 Zielkindern, die von drei Ratergruppen beschrieben wurden: von sich selbst mittels Selbsteinschätzungen, von der Lehrperson sowie von mehreren, zufällig ausgewählten Mitschülerinnen und Mitschülern mittels Fremdeinschätzungen.

Die Basisauswertungen wurden mit SPSS 22 vorgenommen. Zur Bestimmung der Dimensionalität der zentralen Konstrukte bzw. zur Item-Selektion wurden explorative (EFA) und konfirmatorische Faktorenanalysen (CFA) mit Mplus 7.2 berechnet. Ein Ziel dieser Analysen bestand in der Redimensionierung des Instrumentariums, denn bei der Operationalisierung der theoretischen Konstrukte wurde eine große Auswahl möglicher Indikatoren hergeleitet, deren Beantwortung pro Zielkind viel Zeit in Anspruch nimmt. Da bei der geplanten Hauptstudie mehr Zielkinder pro Rater eingeschätzt werden sollen, war es notwendig, die besten Items zu ermitteln und überzählige zu eliminieren.

Bei den Analysen wurden die (teilweise deutlich) rechtsschief verteilten Daten gemäß Empfehlung von Finney und DiStefano (2013) unter Anwendung des verteilungsfreien WLSMV-Schätzverfahrens mittels CATEGORICAL Option analysiert. Da bei den Schülerdaten mehrere individuelle Fremdeinschätzungen (L1) in Zielkindern (L2) verschachtelt sind, wurde mehrebenenanalytisch überprüft, wie konsistent die Mitschüler/innen die Zielkinder eingeschätzt haben und ob die Faktorstruktur ebenenspezifische Unterschiede aufweist.

5 *Instrumentarium*

5.1 *Zielkindspezifische Auftretenshäufigkeit unterrichtlicher Devianz*

In Anlehnung an verfügbare und in Entwicklung befindliche Instrumente (Casale, Hennemann, Huber & Grosche, 2015; Christ, Riley-Tillman & Chafouleas, 2009; Hartke & Vrbán, 2008; Müller, Begert, Gmünder & Huber, 2012; Wettstein, 2008) wurde ein neuer Fragebogen mit 36 Items entwickelt.

Die drei Ratergruppen wurden zum ersten Befragungszeitpunkt (t1) mit angepassten Formulierungen gefragt, wie häufig ein bestimmtes Zielkind in den vergangenen zwei Wochen bestimmte *normabweichende Verhaltensweisen im Unterricht* gezeigt hatte. Die Lehrpersonen schätzten alle sechs Zielkinder ihrer Klasse ein, den Schülerinnen und Schülern wurde nur ein Zielkind zugewiesen. Diese Ratings wurden mittels vierstufigem Antwortformat vorgenommen (0=nie; 1=1 Mal; 2=2 Mal; 3=3 Mal und mehr). In den nachfolgend erörterten Analysen wurden 16 der ursprünglich 36 Items berücksichtigt (vgl. Tab. 1).

Die Item-Selektion erfolgte in mehreren Schritten: Zunächst wurden Items mit einem Mittelwert nahe Null und geringer Varianz ausgeschlossen, z.B. „Beleidigt die Lehrperson“. Diese Verhaltensweisen wurden aufgrund ihres seltenen Auftretens als vernachlässigbar eingeschätzt. Sodann wurde eine Reihe EFAs durchgeführt, um mögliche Dimensionen unterrichtlicher Devianz bzw. die wichtigsten Indikatoren derselben zu ermitteln (die EFAs wurden ohne Messwiederholung am gleichen Datensatz durchgeführt, die Selbsteinschätzungen wurden nicht berücksichtigt). Bei einer ersten Analyse resultierte u.a. ein Faktor, welcher Verhaltensweisen bündelt, die nicht unmittelbar im Unterrichtsgeschehen auftreten, z.B. „Schwänzt die Schule“. Diese Items wurden nach einer theoretischen Priorisierung eliminiert.

Bei nachfolgenden EFAs wurden fortlaufend Items ausgeschlossen, welche zu theoretisch schwer erklärbaren Faktoren mit unterschiedlichen Verhaltensweisen gebündelt wurden oder hohe Nebenladungen aufwiesen. Nach dieser Vorauswahl wurden im Rahmen der CFAs weitere 4 Items ausgeschlossen, die eine zu geringe Ladung aufwiesen oder aufgrund korrelierter Residualvarianzen die Modellgüte beeinträchtigten.

Tabelle 1: Wortlaut der Ratings unterrichtlicher Devianz

Item	Selbsteinschätzung	Fremdeinschätzung durch Mitschüler/innen	Fremdeinschätzung durch die Klassenlehrperson
d1	Ich arbeite nicht an den Aufgaben, die wir i.U. lösen müssen.	Arbeitet nicht an den Aufgaben, die wir lösen müssen.	Beschäftigt sich mit etwas anderem als mit den gestellten Aufgaben.
d3	Ich spreche ohne Aufstrecken i.U., obwohl ich aufstrecken müsste.	Spricht ohne Aufstrecken i.U., obwohl man aufstrecken müsste.	Sagt etwas ohne Aufstrecken, obwohl das verlangt wäre.
d5	Ich rede mit einem a.K. in normaler Lautstärke i.U., obwohl es ruhig sein sollte.	Redet mit einem a.K. in normaler Lautstärke i.U., obwohl es ruhig sein sollte.	Redet in normaler Lautstärke mit einem a.K., obwohl es ruhig sein sollte.
d7	Ich mache Lärm i.U..	Macht Lärm i.U..	Macht unnötigen Lärm.
d12	Ich rede dazwischen, wenn die Lehrperson mir etwas sagt.	Redet dazwischen, wenn die Lehrperson ihm / ihr etwas sagt.	Redet dazwischen, wenn ich ihm/ihr etwas sage.
d13	Ich gehorche nicht sofort, wenn die Lehrperson mir etwas sagt (z.B. mit Schwatzen aufhören), sie muss es mir zweimal sagen.	Gehorcht nicht sofort, wenn die Lehrperson ihm / ihr etwas sagt (z.B. mit Schwatzen aufhören), sie muss es zweimal sagen.	Braucht mindestens zwei Ansagen, bis er/sie einer Ermahnung nachkommt.
d14	Ich rede wütend mit der Lehrperson.	Redet wütend mit der Lehrperson.	Reagiert aufbrausend auf meine Ermahnungen.
d15	Ich gebe der Lehrperson freche Antworten.	Gibt der Lehrperson freche Antworten.	Gibt mir freche Antworten.
d22	Ich lache ein a.K. aus i.U..	Lacht ein a.K. aus i.U..	Lacht ein a.K. aus.
d23	Ich beleidige ein a.K. i.U..	Beleidigt ein a.K. i.U..	Beleidigt ein a.K..
d24	Ich arbeite i.U. nicht mit einem a.K. zusammen, obwohl ich müsste.	Arbeitet i.U. nicht mit einem a.K. zusammen, obwohl er / sie müsste.	Verweigert die Zusammenarbeit mit einem a.K..
d25	Ich rede schlecht über ein a.K. hinter seinem Rücken i.U..	Redet schlecht über ein a.K. hinter seinem Rücken i.U..	Redet schlecht über ein a.K. hinter seinem Rücken.
d29	Ich schmeiße Sachen von einem a.K. auf den Boden i.U..	Schmeißt Sachen von einem a.K. auf den Boden i.U..	Schmeißt Sachen eines a.K.es auf den Boden.
d31	Ich schreie ein a.K. an i.U..	Schreit ein a.K. an i.U..	Schreit ein a.K. an.
d35	Ich schubse ein a.K. i.U..	Schubst ein a.K. i.U..	Schubst ein a.K..
d36	Ich boxe oder trete ein a.K. i.U..	Boxt oder tritt ein a.K. i.U..	Boxt oder tritt ein a.K..

Anmerkungen: Im Beitrag verwendete Abkürzungen (i.U. = im Unterricht; a.K. = anderes/n Kind, Lp = Lehrperson) wurden im Instrument ausgeschrieben

5.2 Subjektives, zielkindspezifisches Störungsempfinden

Um zu erfassen, *wer welche Schüler/innen als wie stark störend empfindet*, wurde ein Instrument mit 10 Items entwickelt, welche mit Bezug auf ein bestimmtes Zielkind mittels

vierstufigen Antwortformats (0=„stimme gar nicht zu“ bis 3=„stimme voll und ganz zu“) zu beantworten sind. Dieses Instrument wurde eine Woche nach der Erfassung der unterrichtlichen Devianz eingesetzt (t2), wobei gleiche Rater-Zielkind-Paare wie zu t1 gebildet wurden und den Schülerinnen und Schülern ein zusätzliches Zielkind zugewiesen wurde. Der zeitliche Abstand verhinderte, dass die Rater ihre Angaben zum Störungsempfinden aktiv ihren Devianz-Ratings angleichen und dadurch die Zusammenhänge artifiziell erhöht worden wären. Die Item-Selektion erfolgte nach dem gleichen Prinzip wie in Kap. 5.1 beschrieben, wobei sich sieben Items als am besten geeignet erwiesen, das schülerspezifische Störungsempfinden zu erfassen (Tab. 2).

Tabelle 2: Wortlaut der Einschätzungen des zielkindspezifischen Störungsempfindens

Item	Einschätzung durch Mitschüler/innen	Einschätzung durch die Klassenlehrperson
s2	...ist für mich ein guter Freund / eine gute Freundin. (r)	...ist bei den anderen Kindern der Klasse beliebt. (r)
s3	...wird von der Lehrperson immer wieder zurecht gewiesen.	...kostet mich viel Energie.
s4	...kann ganze Schulstunden durcheinanderbringen.	...kann ganze Lektionen durcheinanderbringen.
s5	...lenkt mich oft im Unterricht ab.	...hält andere Kinder vom Lernen ab.
s6	...ärgert ständig die Lehrperson.	...ärgert mich ständig.
s7	...ist ein schwieriger Schüler / eine schwierige Schülerin.	...ist verhaltensauffällig.
s10	...habe ich gern. (r)	...habe ich gern. (r)

Anmerkung: Zur Minimierung von Stigmatisierungseffekten wurden möglichst viele Items positiv formuliert, die umgepolt wurden (r), sodass hohe Werte mit einem hohen Störungsempfinden einhergehen.

6 Ergebnisse

6.1 Faktorielle Struktur der theoretischen Konstrukte

Die zwei Konstrukte wurden zunächst separat und nach Ratergruppe getrennt auf der Ebene der individuellen Ratings untersucht, wobei sich mit der definitiven Item-Auswahl anhand

der EFAs mehrere plausible Lösungen mit unterschiedlicher Anzahl Faktoren abzeichneten. Diese Lösungsansätze zur Bestimmung der optimalen Faktorstruktur wurden anschließend mittels CFAs getestet, wobei die Modellgüte der konkurrierenden Modelle miteinander verglichen wurde – wo möglich mittels Chi²-Differenztests.

Danach wurden Modelle berechnet, in denen die beiden Konstrukte gleichzeitig geschätzt wurden, wobei sich eine *dreifaktorielle Struktur* als beste Lösung erwies: Unterrichtliche Devianz wird hierbei als einzelner Faktor modelliert, der alle 16 Devianz-Items bündelt (vgl. Tab. 1). Der zweite Faktor bündelt die Items s3 bis s7 (vgl. Tab. 2), womit die Befragten zum Ausdruck bringen, inwiefern sie die betreffenden Zielkinder als störend bzw. als ablenkend empfinden (kognitives Störungsempfinden). Der dritte Faktor bündelt die Items s2 und s10, mittels derer die Befragten zum Ausdruck bringen, inwiefern sie gefühlsmäßig ihre Beziehung zu den betreffenden Zielkindern beeinträchtigt sehen (affektives Störungsempfinden).

Schließlich wurde *mehrebenenanalytisch* berücksichtigt, dass jedes Zielkind im Durchschnitt von 3.5 Mitschülerinnen bzw. Mitschülern eingeschätzt wurde, indem auf der zweiten Ebene (L2) n=66 aggregierte Werte basierend auf n=234 individuellen Schülerratings (L1) analysiert wurden. Dabei erzielte das dreifaktorielle Mehrebenenmodell eine gute Passung zu den Schülerdaten ($\chi^2=483.84$, df=458, n.s.; RMSEA=.016; CFI=.965). In der Folge wurden die L2-Residualvarianzen auf 0 fixiert und die Faktorladungen über beide Ebenen gleichgesetzt. Diese Maßnahmen implizieren, dass sich die faktorielle Struktur der latenten Variablen mit Bezug auf die aggregierten Zielkind-Ratings (L2) nicht von derjenigen der individuellen Ratings (L1) unterscheidet (Jak, Oort & Dolan, 2014). Dieses restringierte Modell erzielt ebenfalls gute Fit-Werte ($\chi^2=529.71$, df=497, n.s.; RMSEA=.017; CFI=.956) und wird in Abbildung 2 dargestellt.

Die Darstellungen der Strukturgleichungsmodelle (SEM) entsprechen folgender Konvention: Ovale stehen für latente Konstrukte und Kasten für manifeste Indikatoren, deren aggregierte Werte auf L2 in Kreisen abgebildet sind. Auf den Pfeilen zwischen latenten Variablen und Indikatoren sind die standardisierten Faktorladungen notiert (im Schülermodell stehen auf L2 keine Ladungen, da sie mit denen auf L1 gleichgesetzt sind). Grau hinterlegte Ladungen wurden aus Identifikationsgründen auf den Wert 1 fixiert. Die Residuen sind als kleine Pfeile dargestellt. Bidirektionale Pfeile stehen für Korrelationen, deren Koeffizienten auf den Pfeilen notiert sind. Sämtliche Parameter sind – wenn nicht anders vermerkt – auf dem 0.1-Prozent-Alphafehlerniveau signifikant.

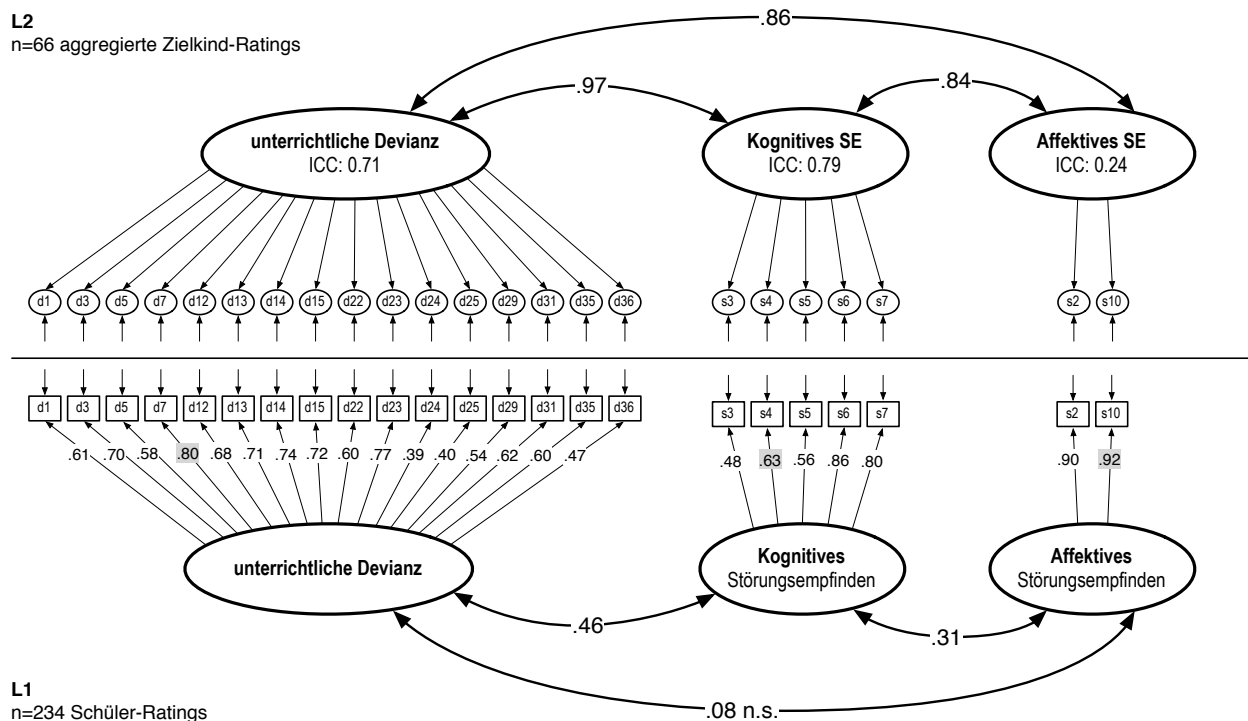


Abbildung 2: Mehrebenenmodell der Schüler-Ratings

Der *Intra-Klassen-Korrelationskoeffizient (ICC)* zeigt an, dass die Schüler/innen sehr konsistent eingeschätzt haben, wie häufig die Zielkinder deviantes Verhalten zeigen: 71% der gesamten Varianz liegen zwischen diesen und widerspiegeln den unter den Ratern geteilten

Varianzanteil. Ähnlich verhält es sich mit dem kognitiven Störungsempfinden: Gemäß dem ICC von .79 werden die Zielkinder von ihren Mitschülerinnen und Mitschülern in konsistenter Weise als (nicht) störend eingeschätzt. Dagegen zeigt sich das zielkindspezifische affektive Störungsempfinden mit einem ICC von nur .24 weit weniger sozial geteilt.

Gemäß der durchschnittlichen Einschätzung der Schüler/innen (L2) werden diejenigen Zielkinder, welche im Unterricht vermehrt deviantes Verhalten zeigen, als störend bzw. ablenkend empfunden ($r=.97$, $p<.001$) sowie weniger gemocht bzw. eher abgelehnt ($r=.86$, $p<.001$). Demgegenüber fallen diese *Zusammenhänge* auf der Ebene der individuellen Ratings (L1) deutlich geringer aus, weil eine beträchtliche Varianz innerhalb der Ratergruppe besteht. Diese ist im Falle des affektiven Störungsempfindens am stärksten ausgeprägt, sodass auf L1 keine signifikante Korrelation zur unterrichtlichen Devianz resultiert.

Weil die CATEGORICAL Option für L2-Variablen für den WLSMV-Schätzer nicht verfügbar ist, konnten die Einschätzungen der Lehrpersonen nicht gemeinsam mit den aggregierten Schüler-Ratings auf der Ebene der Zielkinder (L2) modelliert werden. Entsprechend wurde für die Lehrpersonen in Analogie zu L2 des Schülermodells ein separates Modell spezifiziert, welches ebenfalls gute Fit-Werte erzielt ($\chi^2=238.26$; $df=227$, n.s.; RMSEA=.027; CFI=.998) und in Abbildung 3 dargestellt ist.

Die Korrelationskoeffizienten im Lehrpersonenmodell sind analog zu denjenigen im Schülermodell auf L2 (Abb. 2) zu deuten: Gemäß der Einschätzung der Lehrpersonen werden diejenigen Zielkinder als störend empfunden, welche sich häufig normabweichend verhalten.

Abschließend wurde anhand der manifesten Daten korrelationsanalytisch überprüft, inwiefern Lehrpersonen-Ratings mit den aggregierten Einschätzungen der Schüler/innen (L2)

übereinstimmen. Dabei wurde festgestellt, dass diese zwei Rater-Gruppen das Verhalten der Zielkinder ($r=.86$, $p<.001$, $n=66$) sowie das kognitive Störungsempfinden ($r=.92$, $p<.001$, $n=66$) im Durchschnitt sehr ähnlich einschätzen. Hingegen wird das affektive Störungsempfinden wiederum weniger übereinstimmend eingeschätzt ($r=.43$, $p<.001$, $n=66$).

$n=66$ Lehrpersonen-Ratings

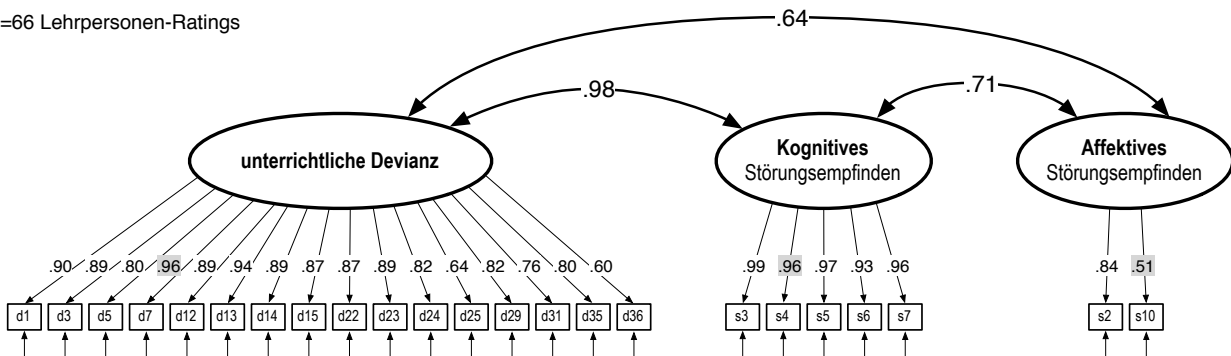


Abbildung 3: Modell der Lehrpersonen-Ratings

6.2 Deskriptive Kennwerte

Der Mittelwert der 16 Devianz-Items über alle Ratings hinweg beträgt 0.39 ($SD=0.48$, $n=517$) und korreliert mit $r=.97$ ($n=516$, $p<.001$) mit dem Mittelwert der ursprünglichen 36 Items ($Mean=0.32$, $SD=0.37$, $n=516$). Normabweichendes Verhalten wird in der untersuchten Stichprobe gesamthaft betrachtet sehr selten gezeigt: Die Selbsteinschätzungen betragen im Mittel 0.31 ($SD=0.31$, $n=63$), die Fremdeinschätzungen der Lehrpersonen 0.46 ($SD=0.59$, $n=66$) bzw. der Mitschüler/innen 0.44 ($SD=0.58$, $n=223$).

Darüber hinaus ist sowohl das kognitive Störungsempfinden von Lehrpersonen ($Mean=0.73$, $SD=0.90$, $n=66$) und von Schülerinnen und Schülern ($Mean=0.76$, $SD=0.83$, $n=445$) als auch das affektive Störungsempfinden von Lehrpersonen ($Mean=0.70$, $SD=0.59$, $n=66$) und Schülerinnen und Schülern ($Mean=1.19$, $SD=0.98$, $n=438$) im Durchschnitt gering ausgeprägt.

Eine nach Rater- und Zielkindgruppe differenzierte Darstellung der Mittelwerte findet sich in Abbildung 4. Sie zeichnet folgendes Bild: Der im Vorfeld der Datenerhebung von den Lehrpersonen subjektiv konstruierte *Kontrast zwischen den drei „Typen“ von Zielkindern* spiegelt sich nicht nur in den Einschätzungen der Lehrpersonen, sondern auch in denen der Mitschüler/innen und der Zielkinder selbst. Auffallend ist, dass die Lehrpersonen diesen Kontrast im Vergleich zu den Schülerinnen und Schülern leicht stärker wahrnehmen.

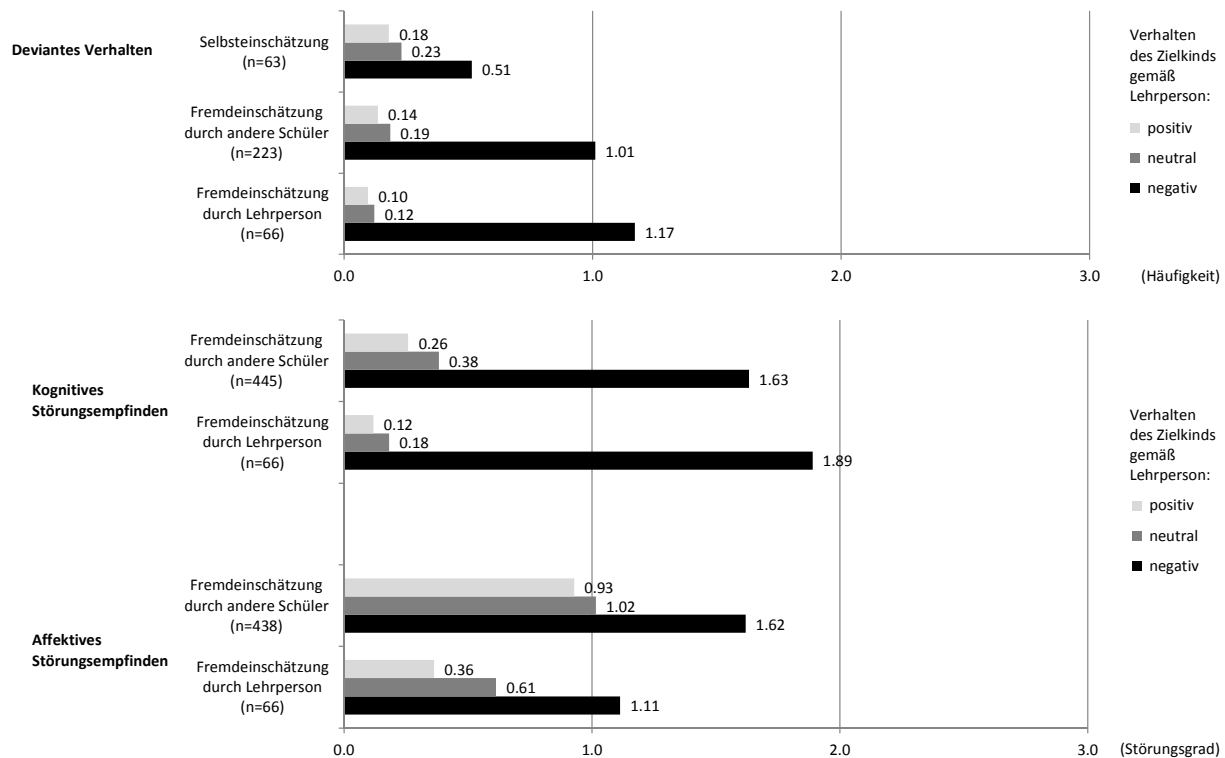


Abbildung 4: Mittelwerte der Konstrukte nach Rater- und Zielkindgruppe

7 Diskussion und Ausblick

Um Unterrichtsstörungen zu erforschen, sollte es das Studiendesign ermöglichen, die theoretische *Entflechtung von Merkmalsdimensionen* empirisch nachzuzeichnen sowie Rater- und Kontexteffekte sichtbar zu machen und möglichst zu kontrollieren. Im Rahmen des SUGUS-Pretests ist das anhand eines Mehrebenenmodells ansatzweise gelungen: Obwohl

diejenigen Zielkinder, welche sich im Unterricht häufiger deviant verhalten, von ihren Mitschülerinnen und Mitschülern durchschnittlich als störender empfunden werden, korreliert das zielkindspezifische Störungsempfinden einzelner Schüler/innen nur teilweise mit ihrer Einschätzung, wie häufig sich die betreffenden Zielkinder normabweichend verhalten. Dieser Befund erhärtet die Annahme, dass das subjektive Störungsempfinden nicht ausschließlich von der Auftretenshäufigkeit unterrichtlicher Devianz bestimmt wird. Damit zeigt sich die Wichtigkeit der systematischen Untersuchung von personalen und kontextuellen Bedingungen beider Konstrukte, was im Rahmen der SUGUS-Hauptstudie geplant ist.

Insgesamt weist das entwickelte *Instrumentarium günstige Kennwerte* auf (z.B. ICCs, Modell-Fits, Signifikanzwerte); dennoch sind gewisse Aspekte kritisch zu betrachten: Items mit Faktorladungen unter .5 sollen ggf. umformuliert, die nach oben gekappten Antwortkategorien zu unterrichtlicher Devianz mit einem halboffenen Antwortformat für größere Häufigkeiten ergänzt werden, und schließlich braucht es weitere, nicht-stigmatisierende Items für das Konstrukt affektives Störungsempfinden.

Eine *Limitation* des Forschungsdesigns betrifft die Erfassung unterrichtlicher Devianz: Trotz des niedrig-inferenten Vorgehens wird nicht tatsächlich gezeigtes, sondern von Ratern wahrgenommenes Verhalten erfasst. Deshalb ist nicht auszuschließen, dass Lehrpersonen bei Zielkindern, die sie als „negativ auffallend“ kategorisiert haben, eher deviantes Verhalten wahrnehmen als bei „positiv auffallenden“ Zielkindern (Hofer, 1986). Eine solche Tendenz findet sich im Ergebnis, dass die Unterschiede zwischen den drei „Typen“ von Zielkindern in den Ratings der Lehrpersonen akzentuierter sind als in den Ratings der Mitschüler/innen und der Kinder selbst.

Darin spiegelt sich eine sampling-spezifische Einschränkung des Pretests: Die Vorauswahl der Zielkinder anhand der subjektiven *Kategorisierung durch die Lehrpersonen* hat möglicherweise deren Rating-Verhalten beeinflusst und zu einer gewissen Verzerrung geführt. Dass diese insgesamt aber begrenzt ist, kommt darin zum Ausdruck, dass die Mittelwerte der Fremdeinschätzungen der Mitschüler/innen und der Lehrpersonen auf vergleichbarem Niveau liegen.

Zur Vermeidung dieses Problems ist beim Einsatz des überarbeiteten Instruments an einer erweiterten, repräsentativen Stichprobe (Zielgröße: 100 Klassen) in der *Hauptstudie* vorgesehen, auf die vorausgehende Kategorisierung von Zielkindern durch die Lehrpersonen zu verzichten. Darüber hinaus ist geplant, mögliche Verzerrungen in der zielkindspezifischen Einschätzung durch die Befragten sichtbar zu machen und zu kontrollieren: Im Rahmen der deskriptiven Analysen durch Differenzentests, im Bereich der SEM stellt der Multitrait-Multimethod Ansatz dazu eine Option dar. Beide Verfahren konnten im Rahmen des Pretests aufgrund der kleinen, nicht repräsentativen Stichprobe nicht angewandt werden, könnten sich bei der Hauptstudie, deren Planung aktuell im Gange ist, dazu als zielführend erweisen.

Für die bevorstehende Datenauswertung stellt die Überprüfung der Faktorstruktur der umschriebenen Konstrukte eine wichtige Aufgabe dar. Das heißt, dass die im vorliegenden Beitrag beschriebenen SEM noch Hypothesencharakter besitzen. Ein besonderes Augenmerk gilt dabei der hohen Korrelation zwischen unterrichtlicher Devianz und kognitivem Störungsempfinden. Von einer Integration der beiden Faktoren wurde bisher aufgrund theoretischer Überlegungen abgesehen. In der Hauptstudie sollte sich diese Frage jedoch klären lassen.

Literatur

- Arbuckle, C. & Little, E. (2004). Teachers' Perceptions and Management of Disruptive Classroom Behaviour during the Middle Years. *Australian Journal of Educational & Developmental Psychology*, 4, 59-70.
- Beaman, R., Wheldall, K. & Kemp, C. (2007). Recent Research on Troublesome Classroom Behaviour. *Australasian Journal of Special Education*, 31(1), 45-60.
- Breidenstein, G. (2006). *Teilnahme am Unterricht*. Wiesbaden: VS.
- Casale, G., Hennemann, T., Huber, C. & Grosche, M. (2015). Testgütekriterien der Verlaufsdiagnostik von Schülerverhalten im Förderschwerpunkt Emotionale und soziale Entwicklung. *Heilpädagogische Forschung*(1), 37-54.
- Christ, T. J., Riley-Tillman, C. & Chafouleas, S. M. (2009). Foundation for the Development and Use of Direct Behavior Rating (DBR) to Assess and Evaluate Student Behavior. *Assessment for Effective Intervention*, 34(4), 201-213.
- Finney, S. J. & DiStefano, C. (2013). Nonnormal and Categorical Data in Structural Equation Modeling. In G. R. Hancock (Hrsg.), *Structural equation modeling a second course* (S. 439-492). Charlotte: Information Age.
- Grosse Siestrup, C. (2010). *Unterrichtsstörungen aus der Sicht von Lehrenden und Lernenden*. Frankfurt: Lang.
- Hartke, B. & Vrbán, R. (2008). *Schwierige Schüler - 49 Handlungsmöglichkeiten bei Verhaltensauffälligkeiten*. Buxtehude: Persen.
- Hillenbrand, C. (2011). *Didaktik bei Unterrichts- und Verhaltensstörungen*. München: Reinhardt.
- Hofer, M. (1986). *Sozialpsychologie erzieherischen Handelns*. Göttingen: Hogrefe.
- Holtappels, H. G. (2000). "Abweichendes Verhalten" und soziale Etikettierungsprozesse in der Schule. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion* (S. 231-256). Opladen: Leske + Budrich.
- Jak, S., Oort, F. J. & Dolan, C. V. (2014). Using Two-Level Factor Analysis to Test for Cluster Bias in Ordinal Data. *Multivariate Behavioral Research*, 49, 544-553.
- Jang, H., Reeve, J. & Deci, E. L. (2010). Engaging Students in Learning Activities. *Journal of Educational Psychology*, 102(3), 588-600.
- Korsch, F. & Petermann, F. (2012). Früherkennung von Verhaltensstörungen durch die schulärztliche Eingangsuntersuchung. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 61(9), 691-705.
- Krause, A., Dorsemagen, C. & Alexander, T. (2011). Belastung und Beanspruchung im Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 788-813). Münster: Waxmann.
- Liesen, C. & Luder, R. (2011). Forschungsstand zur integrativen und separativen schulischen Förderung von Schülerinnen und Schülern mit Verhaltensauffälligkeiten. *Schweizerische Zeitschrift für Heilpädagogik*(9), 11-18.
- Makarova, E., Herzog, W. & Schönbachler, M.-T. (2014). Wahrnehmung und Interpretation von Unterrichtsstörungen aus Schülerperspektive sowie aus Sicht der Lehrpersonen. *Psychologie in Erziehung und Unterricht*, 61(2), 127-140.
- Müller, C. M., Begert, T., Gmünder, L. & Huber, C. (2012). Die Freiburger Selbst- und Peerauskunftsskalen - Schulisches Problemverhalten. *Empirische Sonderpädagogik*(1), 3-21.

- Müller, C. M., Fleischli, J. & Hofmann, V. (2013). *Bericht 1 zur Freiburger Studie zum Peereinfluss in Schulen*. Universität Freiburg.
- Munn, P., Sharp, S., Lloyd, G., Macleod, G., McCluskey, G., Brown, J. et al. (2013). A Comparison of Staff Perceptions of Behaviour in Scottish Schools in 2009 and 2006. *Research Papers in Education*, 28(2), 135-154.
- Mutzeck, W. (2000). *Verhaltensgestörtenpädagogik und Erziehungshilfe*. Bad Heilbrunn: Klinkhardt.
- Myschker, N. & Stein, R. (2014). *Verhaltensstörungen bei Kindern und Jugendlichen*. Stuttgart: Kohlhammer.
- Nickel, H. (1985). Die Lehrer-Schüler-Beziehung aus der Sicht neuerer Forschungsergebnisse. In R. Biermann (Hrsg.), *Interaktion, Unterricht, Schule* (S. 254-280). Darmstadt: Wiss. Buchges.
- Pfützner, M. & Schoppek, W. (2000). Gemeinsamkeiten und Diskrepanzen in der Bewertung von Unterrichtsstörungen durch Lehrer und Schüler. *Unterrichtswissenschaft*, 28(4), 350-378.
- Schönbächler, M.-T., Herzog, W. & Makarova, E. (2011). "Schwierige" Schulklassen: Eine Analyse des Zusammenhangs von Klassenzusammensetzung und wahrgenommenen Unterrichtsstörungen. *Unterrichtswissenschaft*, 39(4), 310-327.
- Schönbächler, M.-T., Makarova, E., Herzog, W., Altin, Ö., Känel, S., Lehmann, V. et al. (2009). *Klassenmanagement und kulturelle Heterogenität. Forschungsbericht Nr. 37*. Universität Bern.
- Schweer, M. K. W. & Thies, B. (2000). Situationswahrnehmung und interpersonales Verhalten im Klassenzimmer. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion* (S. 59-78). Opladen: Leske + Budrich.
- Stein, R. & Stein, A. (2014). *Unterricht bei Verhaltensstörungen*. Bad Heilbrunn: Klinkhardt.
- Textor, A. (2009). Offener Unterricht in der Grundschule mit Schülern mit dem Förderschwerpunkt emotionale und soziale Entwicklung. In C. Röhner, C. Henrichwark & M. Hopf (Hrsg.), *Europäisierung der Bildung* (S. 281-285). Wiesbaden: VS.
- Ulich, K. (1980). *Wenn Schüler stören*. München: Urban + Schwarzenberg.
- Werning, R. (2008). Unaufmerksam und hyperaktiv - der Zappelphilipp in der Schule. In R. Lehberger & U. Sandfuchs (Hrsg.), *Schüler fallen auf*. (S. 116-126). Bad Heilbrunn: Klinkhardt.
- Wettstein, A. (2008). *BASYS. Beobachtungssystem zur Analyse aggressiven Verhaltens in schulischen Settings*. Bern: Huber.
- Wettstein, A. & Thommen, B. (2007). Unterrichtsstörungen stören. *Sonderpädagogik*, 37(3), 156-164.
- Wettstein, A., Thommen, B. & Eggert, A. (2010). Die Bedeutung didaktischer Aspekte in der Aggressionsprävention. *Psychologie in Erziehung und Unterricht*, 57(2), 88-106.
- Zevenbergen, R. (2001). Mathematics, Social Class, and Linguistic Capital. In B. Atweh, H. Forgasz & B. Nebres (Hrsg.), *Sociocultural Research on Mathematics Education* (S. 201-215). Mahwah: Erlbaum.

6.1.3. Unterrichtsstörungen: eine Frage der Perspektive?³³

1. *Ganzheitliches Problemverständnis*

Eine Unterrichtsstörung stellt ein komplexes soziales Phänomen dar, welches sich aus verschiedenen miteinander interagierenden Merkmalsdimensionen konstituiert (Ummel, Wettstein & Thommen, 2009). Um den Problemzusammenhang ganzheitlich zu verstehen, werden in der gegenwärtigen Theoriebildung *verschiedene Erklärungsansätze integriert* – nicht zuletzt, um mittels daraus hervorgehender empirischer Forschung praxisdienliche Bewältigungsstrategien zu identifizieren. Dabei wird u. a. berücksichtigt, dass sich Unterrichtsstörungen immer in einem pädagogisch-didaktischen Kontext ereignen und perspektivisch unterschiedlich wahrgenommen werden können (Stein & Stein, 2014). Auf dieser Suchlinie bewegt sich auch das vom Schweizerischen Nationalfonds geförderte Forschungsprojekt »Studie zur Untersuchung gestörten Unterrichts – SUGUS« (Eckstein, Grob & Reusser, 2016).

2. *Produktion und Rezeption gestörten Unterrichts*

Unterrichtsstörungen können zwar aus verschiedenen Quellen resultieren (Wicki & Kappeler, 2007), doch fokussiert wird häufig das »Problemverhalten« einzelner Schülerinnen und Schüler (Harrison, Vannest, Davis & Reynolds, 2012). Diese Eingrenzung kann dahingehend ergänzt werden, dass nicht alle am Unterricht beteiligten Akteure gleiches Verhalten als problematisch oder störend empfinden. Ein wichtiger Baustein des SUGUS-Projekts besteht in

³³ Copyright © 2018 by Waxmann Verlag. Reproduced [or adapted] with permission. The official citation that should be used in referencing this material is: Eckstein, B. (2018). Unterrichtsstörungen: eine Frage der Perspektive? In S. Schwab, G. Tafner, S. Luttenberger, H. Knauder & M. Reisinger (Hrsg.), *Von der Wissenschaft in die Praxis? Zum Verhältnis von Forschung und Praxis in der Bildungsforschung* (S. 78–92). Münster: Waxmann.

der theoretischen Entflechtung dieses Problemzusammenhangs unter Berücksichtigung des unterrichtlichen Kontexts.

2.1 *Unterrichtliche Devianz*

Viele Schülerinnen und Schüler zeigen im Unterricht gelegentlich *Verhaltensweisen, welche von allgemein oder situativ geltenden Normen abweichen*. Bei dieser unterrichtlichen Devianz handelt es sich meistens um bagatellhafte Formen (z. B. Unkonzentriertheit, motorische Unruhe, Schwatzen), während aggressiv-dissoziales Verhalten gegenüber Mitschülerinnen und Mitschülern sowie oppositionell-trotziges Verhalten gegenüber Lehrpersonen nur selten gezeigt wird (Beaman, Wheldall & Kemp, 2007; Berg, Imhof, Kollera, Schmidt & Ulber, 1998; Carroll, Houghton, Taylor, West & List-Kerz, 2006; Müller, Fleischli & Hofmann, 2013; Munn et al., 2013).

Forschungsarbeiten zur Aktualgenese unterrichtlicher Devianz haben wiederholt auf *präventive Möglichkeiten der Unterrichtsgestaltung* hingewiesen: Da sich die Schülerinnen und Schüler nach eigener Auskunft v. a. bei Über- oder Unterforderung normabweichend verhalten (Schönbächler et al., 2009), können binnendifferenzierende Maßnahmen störungspräventiv wirken (Wettstein, Thommen & Eggert, 2010). Sodann ermöglicht offener Unterricht ein breiteres Spektrum an normkonformen Verhaltensweisen als bspw. Frontalunterricht. Allerdings bedarf auch offener Unterricht einer angemessenen Strukturierung, damit die Schülerinnen und Schüler motiviertes, aufgabenbezogenes Verhalten zeigen (Jang, Reeve & Deci, 2010; Textor, 2009).

2.2 *Subjektives Störungsempfinden*

Unterrichtsstörungen ergeben sich daraus, dass Lehrpersonen oder Schülerinnen und Schüler unterrichtliche Ereignisse mit Bezug auf ihre normativen Erwartungshorizonte als *unangemessen, regelwidrig oder provokativ wahrnehmen* bzw. interpretieren, was sie verärgert, belastet oder vom Lehr-Lernprozess ablenkt (Schweer & Thies, 2000). So werden bereits bagatellhafte Formen normabweichenden Schülerverhaltens von zahlreichen Lehrpersonen als *belastend* erlebt, wenn diese häufig vorkommen (Little, 2005). Im Extremfall kann daraus ein Gesundheitsrisiko für Lehrpersonen entstehen (Kokkinos, 2007; Krause, Dorsemagen & Alexander, 2011). Auch viele Schülerinnen und Schüler empfinden normabweichendes Verhalten ihrer Peers als störend (Infantino & Little, 2005), wobei sie durchschnittlich weniger empfindlich reagieren als die Lehrpersonen (Pfitzner & Schoppek, 2000).

In manchen Klassen unterscheidet sich die Störungswahrnehmung von Lehrpersonen deutlich von derjenigen ihrer Schülerinnen und Schüler (Wettstein, 2013). Solche Befunde deuten darauf hin, dass das Störungsempfinden von anderen Faktoren als vom eigentlichen Verhalten abhängig ist, etwa von personalen Merkmalen der Wahrnehmenden (Lehrperson, Mitschüler/in). Diese *Subjektivität* zeigt sich bspw. auch daran, dass Lehrpersonen mit abnehmender Selbstwirksamkeit Devianz zunehmend als belastend erleben (Arbuckle & Little, 2004; Grosse Siestrup, 2008). Weiter wird das subjektive Störungsempfinden auch von *kontextuellen Faktoren* beeinflusst. Dazu gehören das kollektive Störungsniveau in der Klasse (Makarova, Herzog & Schönbächler, 2014), aber auch didaktische Rahmenbedingungen (Zevenbergen, 2001).

2.3 Interaktionistisches Modell

Diese Befundlage aufgreifend wurde in der SUGUS-Studie in Anlehnung an Nickel (1985) und Wettstein (2012) ein interaktionistisches Theoriemodell der Produktion und Rezeption gestörten Unterrichts entwickelt (Abb. 1). Die *zirkuläre Struktur* des Modells veranschaulicht, dass die aktuelle Störungswahrnehmung auch von früher gefestigten Eindrücken beeinflusst wird (Rosemann & Bielski, 2001) – etwa im Falle von Lehrpersonen, welche gewisse Schülerinnen und Schüler mental als »Störer« eingeordnet haben und sie deshalb in bestimmten Situationen sorgenvoller beobachten als andere (Hofer, 1986).

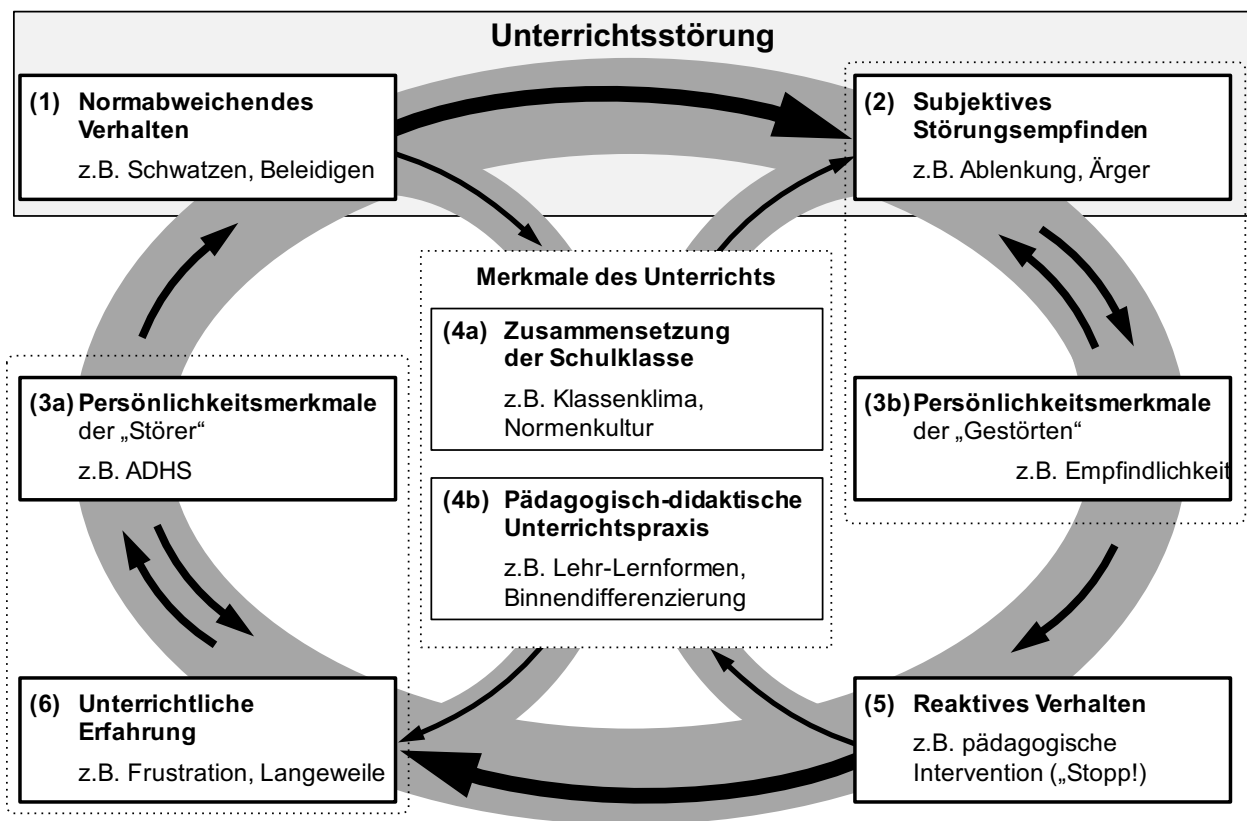


Abbildung 1: Theoriemodell der Produktion und Rezeption gestörten Unterrichts

Das Modell wird im Folgenden anhand eines *beispielhaften, praxisnahen Verlaufs* erklärt:

Zu Beginn der Lektion möchte Lehrerin Muster der ganzen Klasse etwas erklären. Kaum hat sie begonnen, verlässt Eric seinen Platz und tuschelt im Vorbeigehen halblaut etwas zu einem Mitschüler. Erics Verhalten verstößt gegen Normen – vereinbarte Klassenregeln, unausgesprochene Benimmregeln – welche in dieser Unterrichtssituation gelten (1). Frau Muster sowie mehrere Mitschülerinnen und Mitschüler werden aufgrund von Erics Verhalten von der Erklärung abgelenkt und ärgern sich darüber (2).

Solche Szenen ereignen sich mehr oder weniger regelmäßig, denn Eric fällt es aufgrund seiner Veranlagung schwer, lange stillzusitzen (3a). Daher ist Frau Muster ihm gegenüber ziemlich auf Hab-Acht und reagiert manchmal etwas empfindlich auf sein »Gezappel« (3b). Das mag auch daran liegen, dass sein Verhalten im Vergleich zur restlichen, gut funktionierenden Klasse besonders häufig als störend auffällt (4a).

Die aktuelle Situation stört aber schlichtweg auch deshalb, weil sie sich im Klassenunterricht zuträgt (4b). Während einer Gruppenarbeit mit Bewegungs- und Rede-Erlaubnis hätte Erics Verhalten kaum jemanden gestört. Je nachdem, wie Frau Muster nun reagiert (5) und wie Eric diese Reaktion wiederum auffasst (6), kann sich die Situation in eine mehr oder weniger günstige Richtung weiterentwickeln. (Eckstein, Luger, Grob & Reusser, 2016a, S. 2)

3. *Beurteilung der Befundlage und Fragestellung*

Angesichts der im skizzierten Beispielverlauf zutage tretenden Komplexität des Problemzusammenhangs ist die Forschungslage in zweierlei Hinsicht unbefriedigend: Erstens wird den *subjektiven Wahrnehmungsprozessen* forschungsmethodisch häufig nicht adäquat Rechnung getragen, sodass gewisse Befunde vermutlich mit unkontrollierten Ratereffekten konfundiert sind (Korsch & Petermann, 2012). Zweitens fehlen Studien, die auch Kontextvariablen der alltäglichen pädagogisch-didaktischen Unterrichtsgestaltung in die Datenerhebung einbeziehen. Die SUGUS-Studie greift diese Desiderate auf. Im vorliegenden Beitrag werden vorbereitende Basisauswertungen zur späteren Beantwortung der Hauptfragestellungen vorgestellt. Mit den bisherigen Analysen wurden zwei Leitfragen bearbeitet:

- *Gibt es Perspektivenunterschiede in der Wahrnehmung von Unterrichtsstörungen durch Lehrpersonen gegenüber Schülerinnen und Schülern?*
- *Wie können die Befunde den an der Studie teilnehmenden Lehrpersonen kommuniziert werden?*

4. *Forschungsdesign der SUGUS-Studie*

4.1 *Stichprobe*

Im Sommer 2016 wurde in 10 deutschsprachigen Kantonen der Schweiz eine *schriftliche Befragung mit 85 Primarschulklassen* durchgeführt (5. Jahrgangsstufe; inkl. acht jahrgangsgemischte Klassen; Durchschnittsalter: 11,7 Jahre). An der Umfrage beteiligten sich die 85 Klassenlehrpersonen sowie 1.412 aller 1.687 Schülerinnen und Schüler dieser Klassen. Die Befragung fand zu zwei Zeitpunkten jeweils während einer Lektion im Abstand von einer Woche statt. Die Lehrpersonen (Lp) beantworteten mehr Fragen als die Schülerinnen und Schüler (SuS) mit leichten sprachlichen Anpassungen. Dieses Forschungsdesign wurde 2014 im Rahmen einer Pilotstudie erprobt (Eckstein, Reusser, Grob & Hofstetter, 2015) und seither gezielt weiterentwickelt.

4.2 *Instrumentarium*

Alle Befragten – Lehrpersonen und Schüler/innen – erhielten einen personalisierten Fragebogen. Auf einem Abrissstreifen war ihr Name sowie derjenige von zu beschreibenden Zielkindern der Klasse abgedruckt. Es gab einen allgemeinen Teil, u. a. zur pädagogisch-didaktischen Unterrichtsgestaltung, sowie einen spezifischen Teil zu Störungen. Aus Letzterem werden im Folgenden vier Bereiche vorgestellt.

4.2.1 Unterrichtliche Devianz

Zum ersten Befragungszeitpunkt (t1) wurde erfasst, wie häufig einzelne Zielkinder in den vergangenen zwei Wochen *18 normabweichende Verhaltensweisen* zeigten, z. B. »Hat Lärm gemacht im Unterricht«, »Hat mir freche Antworten gegeben«, »Hat im Unterricht ein anderes Kind beleidigt«. Die Lehrpersonen schätzten das Verhalten aller SuS ihrer Klasse ein, die Kinder beschrieben ihr eigenes Verhalten sowie dasjenige von vier zufällig ausgewählten Peers. Das Antwortformat bestand aus sechs Kategorien (»Nie« bis »5 Mal«) plus einer Option für freie Antworten (»häufiger, nämlich:«). Nach Tabachnik und Fidell (2007) wurde nachträglich eine Obergrenze festgelegt, wonach 14 ursprünglich höhere Angaben auf den Wert 60 gekappt wurden (0,44 % aller Selbsteinschätzungen, 0,15 % aller Peer-Ratings). Die Skalenqualität ist gemäß Cronbachs Alpha bei allen drei Ratergruppen gut ($n = 1.677$ Lp-Ratings: $\alpha = 0,92$; $n = 1.307$; Selbsteinschätzungen: $\alpha = 0,86$; $n = 5.082$; Peer-Ratings: $\alpha = 0,91$). Detaillierte Informationen zum Datencleaning und zur Skalierung finden sich im technischen Bericht der SUGUS Studie (Eckstein, Luger, Grob & Reusser, in Vorbereitung).

4.2.2 Subjektives Störungsempfinden

Eine Woche später (t2) beschrieben die Rater die gleichen Zielkinder, deren Verhalten sie zu t1 eingeschätzt hatten. Der Zeitabstand sollte verhindern, dass die Rater ihre Angaben zum Störungsempfinden ihren Einschätzungen zur Devianz bewusst angleichen und damit den Zusammenhang zwischen den beiden Konstrukten artifiziell erhöhen würden. Pro Zielkind wurden *neun Aussagen* mittels vierstufigem Antwortformat beurteilt (»Stimmt gar nicht« bis »Stimmt genau«), z. B. »Hat mich im Unterricht gestört«, »Hat mich geärgert«. Die

Skalenqualität ist wiederum bei allen drei Ratergruppen gut ($n = 1.614$ Lp-Ratings: $\alpha = 0,92$; $n = 5.291$ Peer-Ratings: $\alpha = 0,92$; $n = 1.313$ Selbsteinschätzungen: $\alpha = 0,85$).

Diese beiden zielkindspezifischen Skalen zur Devianz sowie zum Störungsempfinden basieren auf den in der Pilotstudie eingesetzten Instrumenten, deren Qualität mittels konfirmatorischer Faktorenanalyse geprüft wurde (Eckstein et al., 2016). Infolge dieses Pretests wurde das Format der Skalen gezielt weiterentwickelt.

4.2.3 Allgemeine Störungssensitivität

Wie empfindlich die Lehrpersonen und SuS im Allgemeinen auf Störungen reagieren, wurde erfasst, indem sie acht *standardisierte Fallvignetten* beurteilten, z. B. »Zwei Kinder boxen sich im Unterricht gegenseitig unter dem Pult«. Mit vierstufigem Antwortformat beurteilten sie zu t1, ob sie in solchen prototypischen Unterrichtsstörungen genervt sind – zu t2 ob sie dadurch abgelenkt werden (»gar nicht« bis »stark«). Die 16 Items bilden eine akzeptable bis gute Skala ($n = 80$ Lp: $\alpha = 0,79$; $n = 1.250$ SuS: $\alpha = 0,86$).

4.2.4 Allgemeine Belastung

Inwieweit die Lehrpersonen und die SuS den schulischen Alltag als belastend erleben, wurde erfasst, indem sie adressatenspezifische Aussagen anhand eines vierstufigen Antwortformats beurteilten (»Stimmt gar nicht« bis »Stimmt genau«). Die Lehrpersonen beurteilten *sieben Items*, z. B. »Unterrichtsstörungen sind für mich eine große Belastung in meiner aktuellen Klasse«, »Ich fühle mich ausgebrannt« ($n = 81$; $\alpha = 0,76$). Die SuS beurteilten *sechs Items*, z. B. »Lernen im Unterricht ist eine Qual«, »Der Unterricht ist langweilig« ($n = 1.280$; $\alpha = 0,73$).

Auch diese beiden allgemeinen Skalen zur Störungssensitivität und zum Belastungserleben beruhen auf dem im Pretest eingesetzten Instrumentarium (Eckstein et al., 2015) und wurden punktuell weiterentwickelt.

4.3 Auswertungsstrategie

Zunächst sollte ein Überblick über das Datenmaterial erlangt werden, um die Gesamtsituation gemäß den mutmaßlich verschiedenen Perspektiven grob einordnen zu können. Dazu wurden die vier Bereiche *deskriptiv, varianzanalytisch und mittels T-Tests* untersucht. Um die Ergebnisse der zwei wichtigsten Bereiche für die an der Studie teilnehmenden Lehrpersonen in einem für sie erstellten Bericht (Eckstein et al., 2016b) sinnstiftend zu veranschaulichen, wurden sie auf Klassenebene aggregiert und als Balkendiagramme dargestellt. Zudem wurden die Daten *klassenspezifisch ausgewertet und in Netzdiagrammen* abgebildet, um den Lehrpersonen die Situation in ihrer Klasse multiperspektivisch zu illustrieren.

5. Bereichsspezifische Befunde

5.1 Auftretenshäufigkeit unterrichtlicher Devianz

Aus Sicht der Lehrpersonen zeigten die einzelnen SuS *durchschnittlich rund 14 Normabweichungen innerhalb von zwei Wochen* ($\text{Sum} = 14,04$; $SD = 22,53$; $n = 1.660$), hauptsächlich bagatellhafte Formen wie Schwatzen. Bei mittlerer Klassengröße von 19,72 SuS ($SD = 3,08$; $n = 85$) kumulieren sich die individuellen Summenscores auf 277,88 Normabweichungen ($SD = 197,66$; $n = 1.660$), welche sich in der »Durchschnittsklasse« insgesamt ereigneten. Diese mittlere Klassensumme ist in Abbildung 2 im obersten Balken abgebildet (grauer und schwarzer Bereich zusammen). Der schwarze Bereich steht für

durchschnittlich 3,17 SuS pro Klasse, welche ausschließlich von ihrer Lehrperson eingeschätzt wurden.

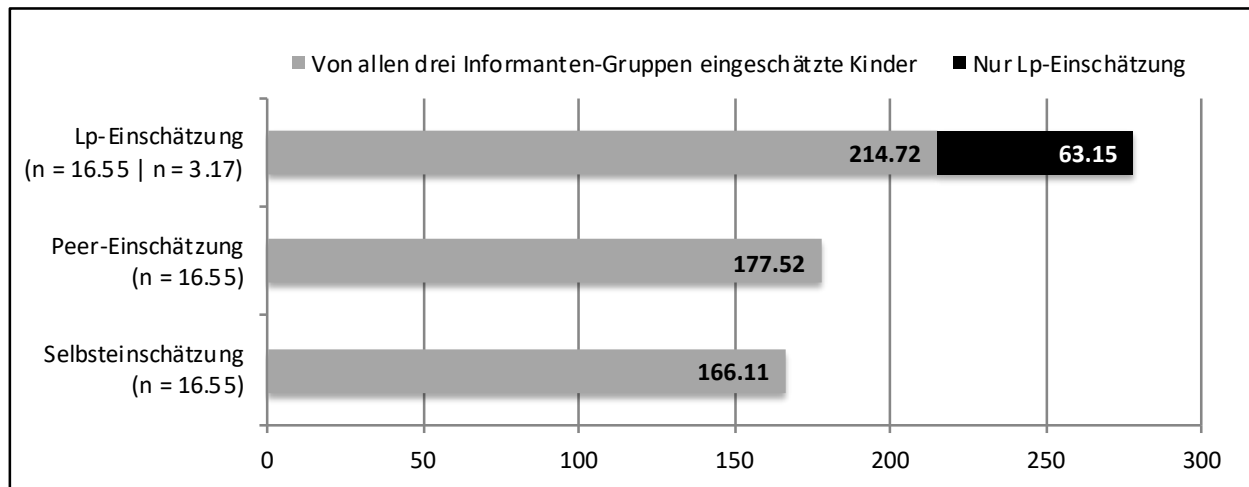


Abbildung 2: Kumulierte Auftretenshäufigkeit unterrichtlicher Devianz in der Durchschnittsklasse nach Ratergruppe

Für den beabsichtigten *Perspektivenvergleich* konnten nur Angaben zu SuS berücksichtigt werden, welche von allen drei Ratergruppen eingeschätzt wurden (graue Balken). Im Mittel sind dies 16,55 SuS, welche aus Sicht der »Durchschnittslehrperson« gemeinsam 214,72 Normabweichungen zeigten ($SD = 155,95$; $n = 1.390$). Demgegenüber fallen die Selbsteinschätzungen (Mean = 166,11; $SD = 79,18$; $n = 1.390$) sowie die Peer-Ratings (Mean = 177,52; $SD = 72,50$; $n = 5.485$) etwas tiefer aus. Dieser Unterschied ist auf Individualebene zwar signifikant, aber von vernachlässigbarer Effektstärke ($n = 1.375$; $F = 19,40$; $df = 1,69$; $p < 0,001$; $\text{Eta}^2_{\text{part}} = 0,01$).

5.2 Ausmaß des subjektiven Störungsempfindens

Den *mittleren Störungsgrad*, welchen die Lehrpersonen pro Schülerin bzw. Schüler empfinden, beläuft sich auf 0,43 ($SD = 0,59$; $n = 1.632$) – ein tiefer Wert angesichts des

theoretischen Skalenmittelwerts von 1,5. Aus den individuellen Summenscores resultiert auf Klassenebene die mittlere Klassensumme von 77,31 ($SD = 46,46$; $n = 1.633$). Dieser Wert bezieht das kumulierte Störungsempfinden der Durchschnittslehrperson bei 19,72 Schülerinnen und Schülern und ist in Abbildung 3 im obersten Balken abgetragen (grauer und schwarzer Bereich zusammen).

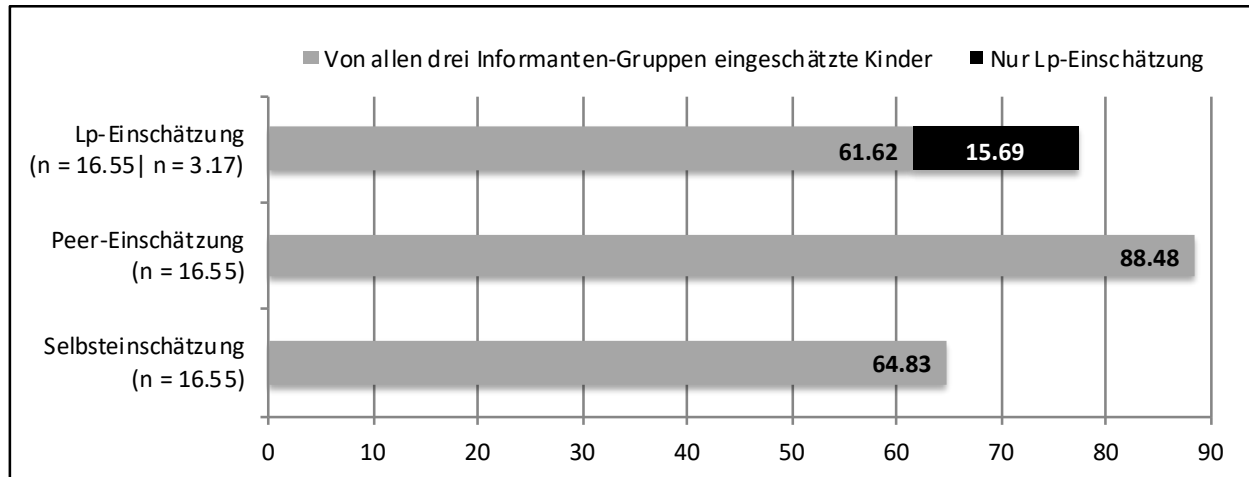


Abbildung 3: Kumulierte Störungsempfinden in der Durchschnittsklasse nach Ratergruppe

Für den *Perspektivenvergleich* wurden erneut nur die Angaben über SuS herangezogen, welche von allen drei Ratergruppen eingeschätzt wurden (graue Balken). Dabei zeigte sich, dass die Lehrpersonen die Zielkinder als ungefähr gleich (wenig) störend einschätzten (Mean = 61,62; $SD = 41,93$; $n = 1.370$) wie die Zielkinder sich selbst (Mean = 64,83; $SD = 26,78$; $n = 1.370$). Demgegenüber beurteilten die Peers ihre Mitschülerinnen und Mitschüler im Durchschnitt als störender (Mean = 88,48; $SD = 29,21$; $n = 5.474$). Dieser Unterschied fällt auf Individualebene bei mittlerer Effektstärke signifikant aus ($n = 1.336$; $F = 96,3$, $df = 1,9$, $p < 0,001$, $\eta^2_{\text{part}} = 0,07$).

5.3 *Störungssensitivität der Lehrpersonen, Schülerinnen und Schüler*

Einige Lehrpersonen und SuS lassen sich schnell ablenken und ärgern sich stark über Unterrichtsstörungen, andere beschreiben sich als viel weniger empfindlich. Diese allgemeine Sensitivität beläuft sich bei den Lehrpersonen auf einen Mittelwert von 1,20 ($SD = 0,42$, $n = 84$), bei den SuS auf 1,26 ($SD = 0,56$; $n = 1.392$). Die beiden Gruppen unterscheiden sich in dieser Hinsicht nicht signifikant voneinander ($t = -1,17$; $df = 101,80$; n.s.).

5.4 *Von Lehrpersonen, Schülerinnen und Schülern erlebte Belastung*

Mit Blick auf die erlebten Belastungen im schulischen Alltag belaufen sich die Angaben der Lehrpersonen auf den Mittelwert von 0,66 ($SD = 0,47$; $n = 83$), bei den SuS auf 0,67 ($SD = 0,50$; $n = 1.355$). Die beiden nahezu identischen Mittelwerte ($t = -0,24$; $df = 1.436$; n.s.) liegen deutlich unterhalb des theoretischen Skalenmittelwerts (1,5), was bedeutet, dass sowohl die Lehrpersonen als auch ihre SuS sich als durchschnittlich wenig belastet erleben.

6. *Zusammenfassender Perspektivenvergleich*

Um die Perspektiven verschiedener Informanten zu den vier zuvor beschriebenen Bereichen vergleichend darzustellen, wurden *Netzdiagramme* erstellt (Eckstein et al., 2016b). Dazu wurde die Skalierung vorgängig mittels linearer Transformation vereinheitlicht, wonach der Wertebereich für alle vier Bereiche von 0 (keine Normabweichung etc.) bis 100 (höchster in der Stichprobe vorkommender Wert) reicht.

6.1 Adressierte Lehrperson vs. Durchschnittslehrperson

In einem ersten, personalisierten Ergebnisbericht wurden die Angaben der adressierten Lehrperson und die mittleren Angaben aller Lehrpersonen in einem Netzdiagramm simultan dargestellt. Bei solchen klassenübergreifenden Vergleichen wurde jeweils darauf hingewiesen, dass aufgrund des Aggregierens in größeren Klassen theoretisch größere Flächen möglich sind als in kleinen, was sich in den Resultaten nicht 1:1 niederschlägt. Abbildung 4 zeigt ein solches Diagramm exemplarisch für die *Beispielklasse »Low 5«*. Diese virtuelle Klasse wurde für eine anonymisierte Fassung des Ergebnisberichts aus denjenigen fünf Klassen ermittelt, bei denen aus den Angaben der Lehrpersonen die kleinste Fläche resultierte (Eckstein et al., 2016a). Die durchschnittlichen Angaben dieser fünf Lehrpersonen sind als weiße Fläche dargestellt, die schwarze Fläche entspricht den durchschnittlichen Angaben aller Lehrpersonen. Das Diagramm veranschaulicht, dass die fünf Lehrpersonen der Klasse »Low 5« ihre Situation durchschnittlich als deutlich weniger problematisch einschätzen als die Durchschnittslehrperson.

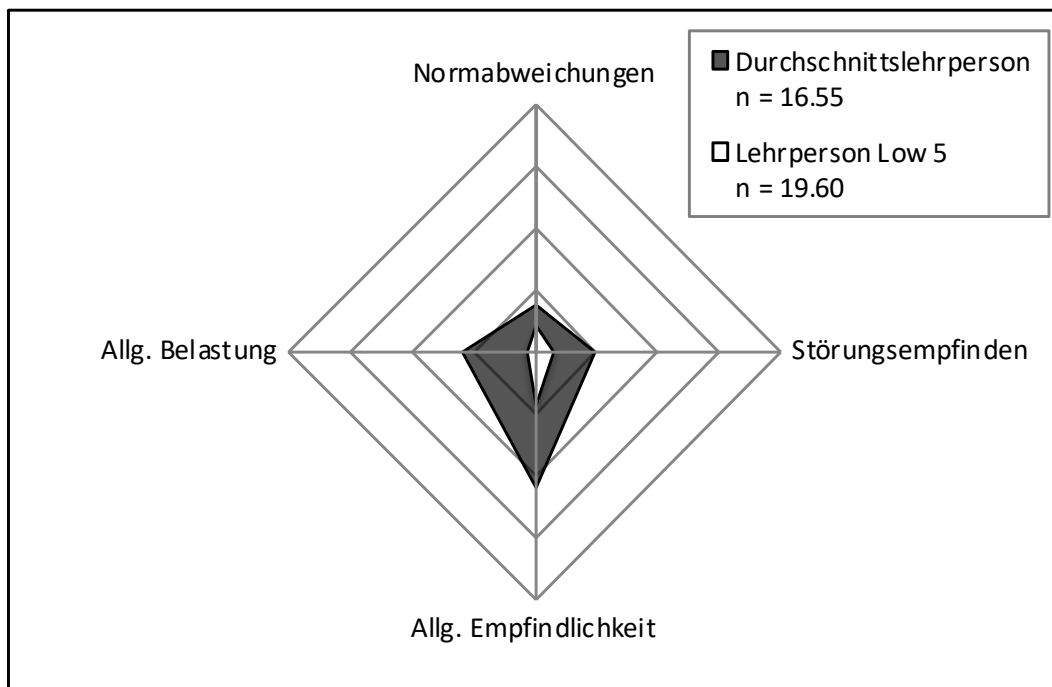


Abbildung 4: Lehrperson der Beispielklasse »Low 5« vs. Durchschnittslehrperson

6.2 Adressierte Lehrperson vs. Schülerinnen und Schüler

In einem anderen Netzdiagramm wurden die Angaben der adressierten Lehrperson mit den mittleren Angaben ihrer Schülerinnen und Schüler vergleichend dargestellt. Abbildung 5 zeigt ein solches Diagramm exemplarisch für die virtuelle *Beispielklasse »Top 5«*, welche aus denjenigen fünf Klassen ermittelt wurde, bei denen aus den Angaben der Lehrpersonen die größte Fläche resultierte (ebd.). Die durchschnittlichen Angaben dieser fünf Lehrpersonen werden als schwarze Fläche dargestellt und mit den durchschnittlichen Angaben der Schülerinnen und Schüler dieser fünf Klassen kontrastiert (weiße Fläche). Das Diagramm zeigt, dass die fünf Lehrpersonen die Situation insgesamt als deutlich problematischer einschätzen als ihre Schülerinnen und Schüler.

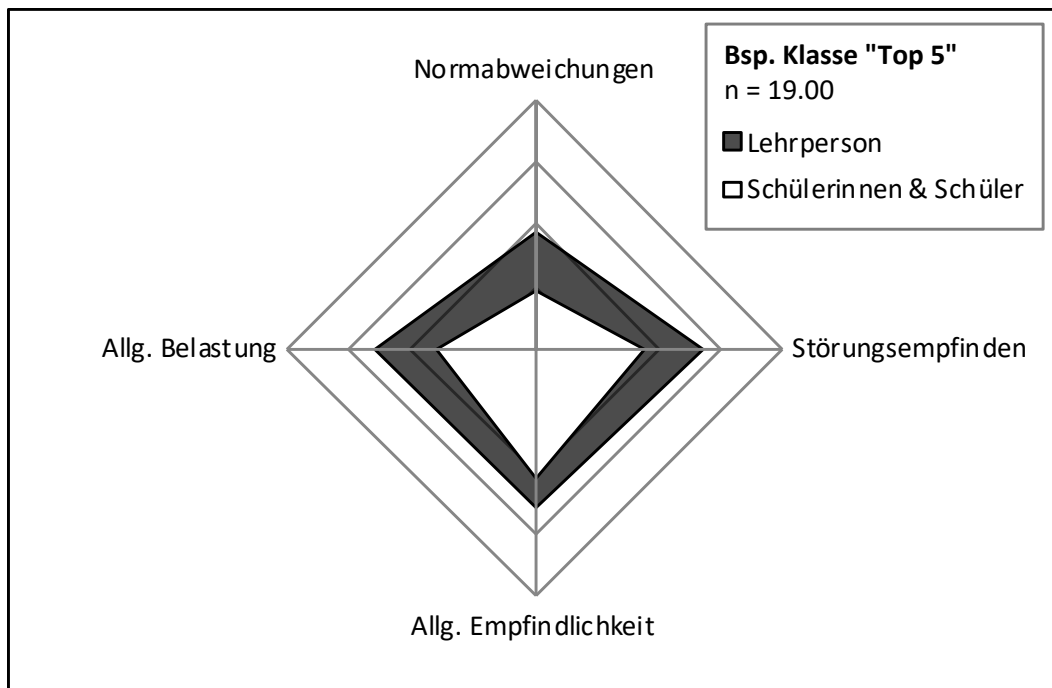


Abbildung 5: Lehrperson vs. Schülerinnen und Schüler der Beispielklasse »Top 5«

7. *Diskussion, Grenzen und Ausblick*

Das SUGUS-Projekt versteht Unterrichtsstörungen als facettenreichen, komplexen Problemzusammenhang, welcher von den beteiligten Akteuren aus ihrer jeweils subjektiven Perspektive wahrgenommen wird. Im vorliegenden Beitrag wurden die Basisauswertungen von vier theoretischen Konstrukten vorgestellt, wovon drei (Devianz, Störungsempfinden, allg. Belastung) in jeder untersuchten Perspektive *rechtsschief* verteilt sind. Angesichts dieser Resultate ist nur in einzelnen Fällen von einer erhöhten Belastung auszugehen. In den meisten Klassen bewegt sich die berichtete Störungs- und Belastungsintensität in einem Ausmaß, welches vermutlich gut zu bewältigen ist.

Hinsichtlich der interessierenden Perspektivenunterschiede lassen die auf der Stichprobenebene ermittelten nicht signifikanten Resultate bzw. geringen Effektstärken noch keine endgültigen Schlussfolgerungen zu. Dahinter werden Grenzen des Untersuchungssettings vermutet: Erstens konnten aus forschungsethischen Gründen nur diejenigen SuS befragt werden, die sich mit dem Einverständnis ihrer Eltern *freiwillig* an der Studie beteiligten. Insgesamt 275 Kinder wollten oder durften keinen Fragebogen ausfüllen, weshalb auch davon abgesehen wurde, sie als Zielkinder von ihren Peers beschreiben zu lassen. Als Korrektiv wurden jedoch Ratings der Lehrpersonen über sie eingeholt, wonach just diese SuS häufiger ($t = -3,6$; $df = 314,2$; $p < 0,001$) deviantes Verhalten zeigen als diejenigen, welche sich aktiv an der Umfrage beteiligten. Es könnte sein, dass gerade die nicht teilnehmenden SuS divergent wahrgenommen werden, wonach die ermittelten Befunde die tatsächlichen Perspektivenunterschiede unterschätzen. Zweitens ist zu vermuten, dass die Resultate deutlicher ausfallen werden, wenn die Konstrukte anhand faktorenanalytisch bestimmter Subdimensionen mehrbenenanalytisch untersucht werden – diese Analysen stehen noch aus. Ungeachtet dieser Limitationen wurden auf

der Ebene einzelner Schulklassen teilweise erhebliche Perspektivenunterschiede zwischen Lehrpersonen und SuS ermittelt, auf sie auch Wettstein, Ramseier, Scherzinger und Gasser (2016) nachweisen konnten.

Im vorliegenden Beitrag wurde dargelegt, wie diese Befunde den an der Studie teilnehmenden Lehrpersonen in Form eines Ergebnisberichtes kommuniziert worden sind (Eckstein et al., 2016b). Auf diesen Resultaten basiert auch die weiterführende *Auswertungsstrategie im SUGUS-Projekt*, deren Stoßrichtung folgende Korrelationsanalysen aufzeigen: Basierend auf den mittleren Einschätzungen mehrerer SuS pro Zielkind (Aggregation auf Ebene Zielkind) zeigt sich ein starker Zusammenhang zwischen deviantem Verhalten und Störungsempfinden ($r = 0,71$). Das bedeutet: Je häufiger sich ein Zielkind normabweichend verhält, desto mehr stört es seine Peers – ein wenig erstaunlicher Befund. Jedoch fällt die Korrelation bei nicht aggregierten Peer-Ratings mit $r = 0,54$ tiefer aus ($z = 9,45$; $p < 0,001$). Demnach unterliegt das Störungsempfinden subjektiven Einflüssen, wie auch weitere Analysen andeuten: Das mittlere Störungsempfinden der SuS bezüglich der von ihnen beurteilten Zielkinder (Aggregation auf Ebene Rater) korreliert mit $r = 0,18$ mit ihrer allgemeinen Störungssensitivität. Bei den Lehrpersonen ist dieser Zusammenhang mit $r = 0,40$ stärker ($z = -2,10$; $p < 0,05$). Das bedeutet: Je empfindlicher die Befragten sind, desto mehr fühlen sie sich von einzelnen Kindern gestört – unabhängig vom Verhalten dieser Kinder.

In den geplanten weiterführenden Analysen stehen neben solchen subjektiven Einflüssen auch Kontextfaktoren im Fokus. Dabei soll u. a. untersucht werden, ob die alltägliche pädagogisch-didaktische Unterrichtsgestaltung die Auftretenshäufigkeit normabweichenden Verhaltens oder das Ausmaß des subjektiven Störungsempfindens beeinflusst.

Literatur

- Arbuckle, C. & Little, E. (2004). Teachers' Perceptions and Management of Disruptive Classroom Behaviour During the Middle Years. *Australian Journal of Educational & Developmental Psychology*, 4, 59–70.
- Beaman, R., Wheldall, K. & Kemp, C. (2007). Recent Research on Troublesome Classroom Behaviour: A Review. *Australasian Journal of Special Education*, 31(1), 45–60.
- Berg, D., Imhof, M., Kollera, S., Schmidt, U. & Ulber, D. (1998). Häufigkeiten von Verhaltensauffälligkeiten in der Grundschule aus der Sicht der Klassenlehrer. *Psychologie in Erziehung und Unterricht*, 45, 280–290.
- Carroll, A., Houghton, S., Taylor, M., West, J. & List-Kerz, M. (2006). Responses to Interpersonal and Physically Provoking Situations. *Educational Psychology*, 26(4), 483–498.
- Eckstein, B., Grob, U. & Reusser, K. (2016). Unterrichtliche Devianz und subjektives Störungsempfinden. Entwicklung eines Instrumentariums zur Erfassung von Unterrichtsstörungen. *Empirische Pädagogik*, 30(1), 113–129.
- Eckstein, B., Luger, S., Grob, U. & Reusser, K. (2016a). *SUGUS – Studie zur Untersuchung gestörten Unterrichts. Ergebnisbericht der Hauptstudie – anonymisierte Fassung*. Universität Zürich. Verfügbar über <http://www.ife.uzh.ch/SUGUS> [14.2.2017].
- Eckstein, B., Luger, S., Grob, U. & Reusser, K. (2016b). *SUGUS – Studie zur Untersuchung gestörten Unterrichts. Ergebnisbericht der Hauptstudie in 85 unveröffentlichten, personalisierten Fassungen*. Universität Zürich.
- Eckstein, B., Luger, S., Grob, U. & Reusser, K. (in Vorbereitung). *SUGUS: Technischer Bericht der quantitativen Teilstudie. Studiendesign, Stichprobe und Skalendokumentation*.
- Eckstein, B., Reusser, K., Grob, U. & Hofstetter, A. (2015). *SUGUS – Studie zur Untersuchung gestörten Unterrichts. Kurzer Ergebnisbericht der Vorstudie in elf unveröffentlichten, personalisierten Fassungen*. Universität Zürich.
- Grosse Siestrup, C. (2008). *Unterrichtsstörungen aus der Sicht von Lehrenden und Lernenden*. Frankfurt a. M.: Peter Lang.
- Harrison, J., Vannest, K., Davis, J. & Reynolds, C. (2012). Common Problem Behaviors of Children and Adolescents in General Education Classrooms in the United States. *Journal of Emotional and Behavioral Disorders*, 20(1), 55–64.
- Hofer, M. (1986). *Sozialpsychologie erzieherischen Handelns*. Göttingen: Hogrefe.
- Infantino, J. & Little, E. (2005). Students' Perceptions of Classroom Behaviour Problems and the Effectiveness of Different Disciplinary Methods. *Educational Psychology*, 25(5), 491–508.
- Jang, H., Reeve, J. & Deci, E. L. (2010). Engaging Students in Learning Activities. *Journal of Educational Psychology*, 102(3), 588–600.
- Kokkinos, C. M. (2007). Job Stressors, Personality and Burnout in Primary School Teachers. *British Journal of Educational Psychology*, 77, 229–243.
- Korsch, F. & Petermann, F. (2012). Früherkennung von Verhaltensstörungen durch die schulärztliche Eingangsuntersuchung. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 61(9), 691–705.
- Krause, A., Dorsemagen, C. & Alexander, T. (2011). Belastung und Beanspruchung im Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (S. 788–813). Münster: Waxmann.

- Little, E. (2005). Secondary School Teachers' Perceptions of Students' Problem Behaviours. *Educational Psychology*, 25(4), 369–377.
- Makarova, E., Herzog, W. & Schönbächler, M.-T. (2014). Wahrnehmung und Interpretation von Unterrichtsstörungen aus Schülerperspektive sowie aus Sicht der Lehrpersonen. *Psychologie in Erziehung und Unterricht*, 61(2), 127–140.
- Müller, C. M., Fleischli, J. & Hofmann, V. (2013). *Verhaltensprobleme von Jugendlichen auf der Sekundarstufe I. Die Situation im ersten Schuljahr (7. Klasse). Bericht 1 zur Freiburger Studie zum Peereinfluss in Schulen (FRI-PEERS)*. Fribourg: Universität Fribourg (CH).
- Munn, P., Sharp, S., Lloyd, G., Macleod, G., McCluskey, G., Brown, J. & Hamilton, L. (2013). A Comparison of Staff Perceptions of Behaviour in Scottish Schools in 2009 and 2006. *Research Papers in Education*, 28(2), 135–154.
- Nickel, H. (1985). Die Lehrer-Schüler-Beziehung aus der Sicht neuerer Forschungsergebnisse. In R. Bierman (Ed.), *Interaktion-Unterricht-Schule* (S. 254–280). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Pfützner, M. & Schoppek, W. (2000). Gemeinsamkeiten und Diskrepanzen in der Bewertung von Unterrichtsstörungen durch Lehrer und Schüler. *Unterrichtswissenschaft*, 28(4), 350–378.
- Rosemann, B. & Bielski, S. (2001). *Einführung in die pädagogische Psychologie*. Weinheim: Beltz.
- Schönbächler, M.-T., Makarova, E., Herzog, W., Altin, Ö., Känel, S., Lehmann, V. & Milojevic, S. (2009). *Klassenmanagement und kulturelle Heterogenität: Ergebnisse 2*. Universität Bern.
- Schweer, M. & Thies, B. (2000). Situationswahrnehmung und interpersonales Verhalten im Klassenzimmer. In M. Schweer (Ed.), *Lehrer-Schüler-Interaktion* (S. 59–78). Opladen: Leske + Budrich.
- Stein, R. & Stein, A. (2014). *Unterricht bei Verhaltensstörungen*. Bad Heilbrunn: Klinkhardt.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Pearson u. a.
- Textor, A. (2009). Offener Unterricht in der Grundschule mit Schülern mit dem Förderschwerpunkt emotionale und soziale Entwicklung. In C. Röhner, C. Henrichwark & M. Hopf (Hrsg.), *Europäisierung der Bildung* (S. 281–285). Wiesbaden: VS.
- Ummel, H., Wettstein, A. & Thommen, B. (2009). Der verhinderte Unterricht. *Empirische Sonderpädagogik*, 1(1), 80–95.
- Wettstein, A. (2012). A Conceptual Frame Model for the Analysis of Aggression in Social Interactions. *Journal of Social, Evolutionary, and Cultural Psychology*, 6(2), 141–157.
- Wettstein, A. (2013). Die Wahrnehmung sozialer Prozesse im Unterricht. *Schweizerische Zeitschrift für Heilpädagogik*, 19(7/8), 5–13.
- Wettstein, A., Ramseier, E., Scherzinger, M. & Gasser, L. (2016). Unterrichtsstörungen aus Lehrer- und Schülersicht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48(4), 171–183.
- Wettstein, A., Thommen, B. & Eggert, A. (2010). Die Bedeutung didaktischer Aspekte in der Aggressionsprävention. *Psychologie in Erziehung und Unterricht*, 57(2), 88–106.
- Wicki, W. & Kappeler, S. (2007). *Beobachtete Unterrichtsstörungen bei erfahrenen Lehrpersonen im Spiegel subjektiver Ursachenzuschreibungen* (Manuskript). Verfügbar über <http://www.dev.phlu.ch> (18.12.2017).
- Zevenbergen, R. (2001). Mathematics, Social Class, and Linguistic Capital. In B. Atweh, H. Forgasz & B. Nebres (Eds.), *Sociocultural Research on Mathematics Education* (pp. 201–215). Erlbaum.

6.1.4. Production and Perception of Classroom Disturbances³⁴

Abstract

Classroom disturbances impair the quality of teaching and learning, and they can be a source of strain for both teachers and students. Some studies indicate, however, that not everyone involved gets equally disturbed by the same occurrences. Altogether, there is still little solid knowledge about the teachers' and the students' subjective perception of disturbance. Moreover, rater effects may have confounded the findings available. Addressing these desiderata, the X study investigates two elements of classroom disturbances within an interactionist framework: the incidence of deviant behaviour shown by particular target students, and the intensity of disturbance as subjectively perceived by teachers, by classmates, and by the targets themselves. For this purpose, we conducted a questionnaire survey among 85 primary-school class teachers and 1412 students. The data were analysed by means of a two-level correlated trait – correlated method minus one [CT-C(M-1)] model. This relatively novel statistical procedure has only rarely been applied in educational research so far. It made it possible to determine the respondents' common view on classroom disturbances as well as the rater-specific perspectives. The results indicate that increasing deviance coincides with increasing distraction and annoyance – but mainly in a relatively small intersection of the different perspectives. Beyond that, the analysis revealed substantial rater effects which explain 30 to 61% of variance in teacher ratings, for instance. The author discusses likely reasons why disturbances are perceived so divergently.

³⁴ Eckstein, B. (under review). Production and Perception of Classroom Disturbances – A new approach to investigating the perspectives of teachers and students.
[The manuscript has been anonymised for the purpose of blind review]

1 Two main elements of classroom disturbances

Classroom disturbances arise from inappropriate student or teacher behaviour (Montuoro & Lewis, 2015), albeit not everyone in class gets equally distracted or annoyed (Authors, 2016a). This implies that classroom disturbances consist of an objective core that the persons involved may perceive differently (Author, 2018). This argument is theoretically well-founded, but there is only little empirical evidence regarding commonalities and differences between the teachers' and the students' distinct perceptions. This paper presents a new methodological approach to investigating three perspectives on the objective core of classroom disturbances: the self-perception of students, the perception of teachers, and the perception of classmates. Building on two main lines of theory and research, the focus lies on deviant student behaviour and on the persons' subjective perception of disturbance.

1.1 Deviant student behaviour

A first line of theory and research on classroom disturbances has focused on student behaviour that deviates from specific classroom rules (e.g. chattering) or from common socio-moral conventions (e.g. insolence). Studies on incidence rates from several countries largely agree that the most frequent forms of deviant student behaviour are relatively minor *discipline problems* whereas aggressive and *dissocial behaviours* are considerably rarer (Beaman, Wheldall, & Kemp, 2007; Crawshaw, 2015). Many studies investigated ontogenetically determined risk factors of deviant student behaviour, e.g. impulsivity (Carroll, Houghton, Taylor, West, & List-Kerz, 2006). Other studies examined proximal causes and preventions like teaching styles (Godwin et al., 2016; Sherman, Rasmussen, & Baydala, 2008), or classroom management (Emmer & Sabornie, 2015).

This body of research has contributed important knowledge of one aspect of classroom disturbances. However, not all studies considered that the same behaviours can be differently perceived, interpreted, and judged from distinct perspectives (Crawshaw, 2015).

1.2 Teachers' and students' subjective perception of disturbance

A second line of theory and research has revealed differential perceptions of classroom disturbances: most teachers feel stressed when their students frequently show deviant behaviours (E. Little, 2005), which may even result in burnout (Kokkinos, 2007). Many students perceive deviant behaviour of their classmates as disturbing too (Infantino & Little, 2005), but they usually deem it less troubling than teachers (Montuoro & Lewis, 2015). While most students admit that they get *distracted* by their classmates' deviance, not all of them claim to feel *annoyed* (Schönbächler et al., 2009). Students who behave deviantly themselves often realise that this may disturb the others, but they worry much more about their own problems (Preuss-Lausitz, 2005). Altogether, this means that teachers and students perceive classroom disturbances differently according to distinct frames of perception (Wettstein, Ramseier, Scherzinger, & Gasser, 2016).

Role-specific traits influence the teachers' and the students' perception: teachers are accountable for all classroom interactions whereas the students are mainly concerned with their own learning, motivation and emotional wellbeing. This implies distinct valences and normative expectations that affect the way in which teachers and students perceive disturbances (Wettstein, Scherzinger, & Ramseier, 2018). Moreover, *individual traits* of students (Wettstein, Ramseier, & Scherzinger, 2018) and teachers (Hamre, Pianta, Downer, & Mashburn, 2008) influence their perception: teachers with decreasing self-efficacy beliefs, for instance, feel increasingly strained

by deviant student behaviour (Arbuckle & Little, 2004; Dicke et al., 2014). Finally, *contextual traits* affect the perception: In instructional settings with rigid rules (Zevenbergen, 2001) or in classes with a low collective level of disturbances (Makarova, Herzog, & Schönbächler, 2014), occasional occurrences of deviance are perceived as strongly disturbing.

2 *Holistic conceptualisation of classroom disturbances in the X study*

The X study integrates the above introduced two lines of theory and research. We conceptualise classroom disturbances as a co-constructed, *interactionist phenomenon* (Authors, 2016b). Figure 1 illustrates this holistic understanding in a theoretical model (Authors, 2016a; after Wettstein, 2012). The arrows symbolise causal effects. The key argument is depicted in the upper part of the figure: classroom disturbances are constituted by two main elements – deviant student behaviour, and subjective perception of disturbance. We assume that deviance commonly disturbs teachers and students, yet the context as well as role-specific and individual traits of the “disturbed” affect the intensity in which they get distracted and/or annoyed. The present article focuses on this presumed core mechanism.

Beyond that, the model outlines how the assumed interaction could continue: teachers or students might react to behaviours that they have perceived as disturbing, e.g. with rebukes. The “disturber”, in turn, deems such behavioural reactions more or less fair or humiliating – depending on the context and his/her personal traits. The model’s circular structure visualises that student behaviour and the way in which it is perceived, interpreted, and judged are contextualised in a history of preceding interactions (Doyle, 2006).

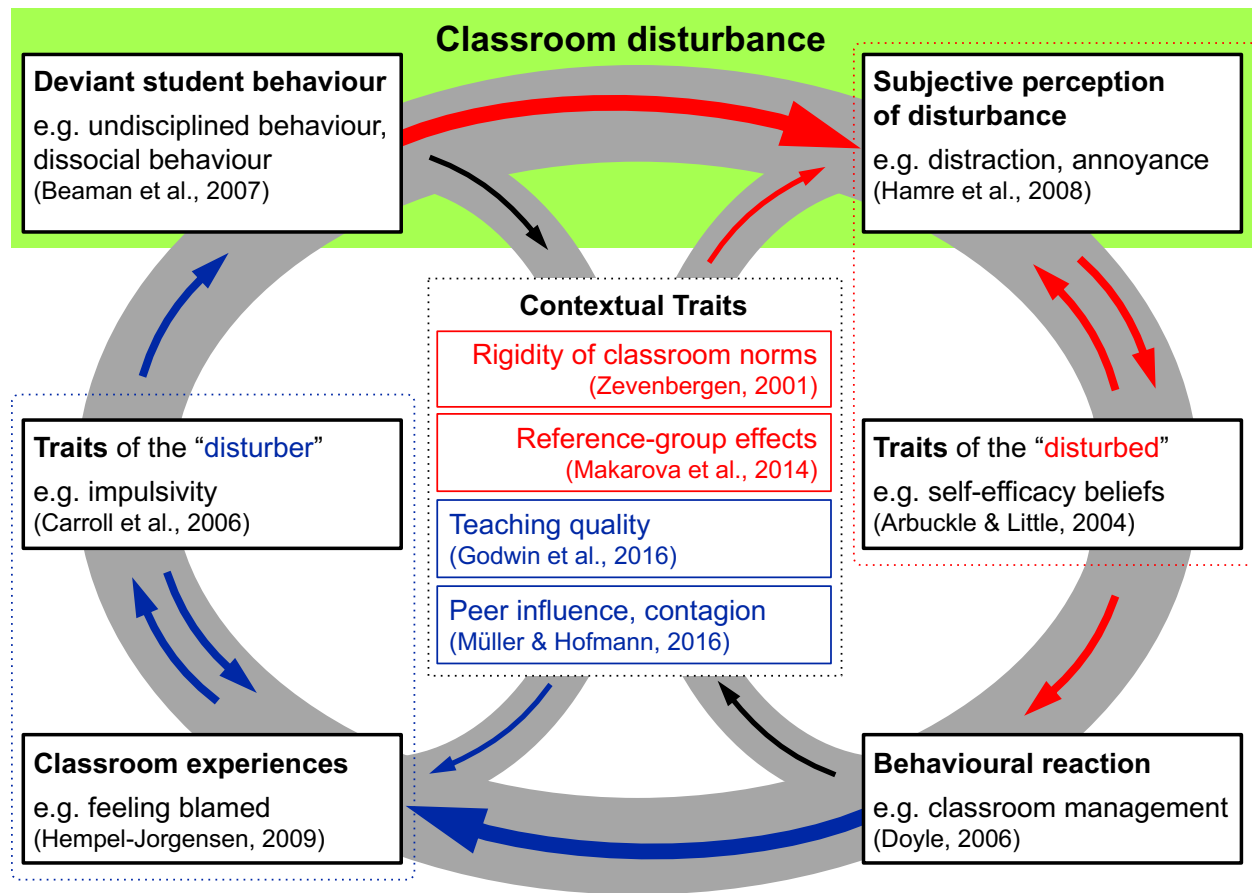


Figure 1. Interactionist model of the production and perception of classroom disturbances.

3 Rater effects as a methodological consequence of differential perceptions

Rater effects stem from characteristics of the raters, the rating instrument, and the rating environment (Wolfe, 2004). In research on classroom disturbances, the above mentioned *distinct frames of perception* almost inevitably entail such rater effects: teachers, students, or external observers perceive and thus assess disturbances according to role-specific, individual and/or contextual conditions. Not all studies that focused on deviant student behaviour accounted for these issues sufficiently (Crawshaw, 2015). As a consequence, some findings available cannot be considered as purely objective information.

Extreme cases are *rater biases*, caused, for example, by halo effects (Hoyt, 2000), prejudiced, selective attention (Hofer, 1986), expectations triggered by labels like “ADHD” (Ohan, Visser, Strain, & Allen, 2011), or self-serving strategies (Fishbein & Ajzen, 2010). Further rater effects may be caused by differential opportunities in terms of the perceptibility of certain disturbances: if students behave deviantly during group work outside the classroom, for instance, only the present classmates will notice but not the teacher.

Moreover, *high-inference rating instruments* that require subjective interpretations from the raters amplify rater effects (Hoyt & Kerns, 1999; Südkamp, Kaiser, & Möller, 2012). Examples from research on classroom disturbances are vague formulations like “troublesome behaviours”, Likert-type frequency scales without clearly defined options (e.g. “seldom – sometimes – often”), or ratings of the whole class instead of individual students.

4 *Aims*

4.1 *General objectives of the research design*

We designed the X study in order to investigate classroom disturbances according to our holistic conceptualisation; and we aimed to control for rater effects (Authors, 2016a). For these purposes, we conducted a multi-perspective survey in which teachers and students reported on individual students of their class (target students): the teachers described all students in their class, the students described themselves plus four randomly assigned classmates (Authors, 2018).

A first objective of the X study was to measure the *incidence of deviant student behaviour* as unbiased as possible: the respondents assessed the targets’ behaviours on a low-inference rating scale that we had developed in order to reduce the amount of potential rater effects. A second objective was to measure the respondents’ *subjective perception of disturbance*:

the respondents assessed the intensity of distraction and annoyance the targets had caused according to their perspectives (Authors, 2016a). A third objective has been to analyse the *relationship between deviance and perception of disturbance* in each perspective and in the raters' common view respectively. By accomplishing this comprehensive research design, the X study breaks new ground in investigating the production and the perception of classroom disturbances.

4.2 *Analysis strategy, research questions, and hypotheses*

The survey yielded a complex multitrait-multimethod data set which has been analysed by means of the “*Correlated Trait – Correlated Method Minus One*” [CT-C(M-1)] approach (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). This is a special variant of structural equation modelling which analyses the influence of latent traits on manifest indicators as well as the impact of the methods applied, e.g. different types of rating.

A CT-C(M-1) model estimates trait factors and method factors. A *trait factor* comprises the amount of the measurements' variance that is consistent across all methods applied. A *method factor*, by contrast, comprises the amount of variance which is specific to this method. One method is selected as comparison standard (reference method) for which no method factor is modelled. As a consequence, there is one method factor less than methods applied, hence “M-1” (Eid et al., 2008).

Psychologists developed this modelling technique originally and applied it, for instance, in personality research with multirater designs in order to analyse the *consistency and the specificity* of different perspectives on the mood of target persons (Carretero-Dios, Eid, & Ruch, 2011). Educational researchers applied this approach only recently in a few studies to analyse

commonalities and differences between teacher and student ratings, e.g. regarding inclusion (Venetz, Zurbriggen, & Schwab, 2017).

The CT-C(M-1) model in this paper analyses teacher, peer and self-ratings of classroom disturbances. The model's trait factors are the intersection of the three perspectives and represent, thus, the raters' common view. In order to quantify the extent of this common view, *consistency* coefficients are calculated. As opposed to this, the model's method factors comprise method-specific divergences from the common view. These divergences can be interpreted as rater effects which are quantified by *specificity* coefficients.

Furthermore, the model analyses correlations. The correlations between the trait factors estimate the *strength of the constructs' relation* in the common view. They indicate how strongly deviant behaviour correlates with the perceived intensity of disturbance – in the shared perspective of all raters. The correlations between the method factors, by contrast, estimate the similarity of rater-specificities across different constructs. They represent the *generalisability of rater effects*.

The analyses provide answers to the following research questions:

(Q-1) *Consistency of deviance ratings*: To what extent are teacher, peer and self-ratings of deviant student behaviour consistent, and to what extent are they rater-specific? Because the ratings rest on a low-inference instrument, a larger extent of consistency than specificity is expected (Hypothesis 1).

(Q-2) *Specificity of the subjective perception of disturbance*: To what extent are teacher, peer and self-ratings of the intensity of disturbance consistent, and to what extent are they rater-specific? Because perception primarily concerns the raters, a larger extent of specificity than consistency is expected (Hypothesis 2).

(Q-3) Relation of deviance and perception of disturbance in the raters' common view:

How strongly do the trait factors of deviance and perception of disturbance correlate? Medium to strong correlations are expected: increasing deviance coincides with increasing distraction and annoyance (Hypothesis 3).

(Q-4) Generalisability of rater-effects: How strongly do the method factors correlate?

Medium correlations are expected: a rater-specific divergence in one construct coincides with analogous divergences in the other constructs (Hypothesis 4).

5 Method

5.1 Sample and conduct of the survey

In summer 2016, we conducted a *written survey* in 85 primary school classes (90.2% fifth grade, 9.8% mixed grades; students' mean age: 11.73 years, SD: 0.52 years). All 85 class teachers and 1412 students out of a total of 1687 participated in the study. 275 students did not participate because they did not want to or because their parents did not allow it. The survey took place twice with one week in between and lasted a whole lesson each time. This research design had been pretested in a pilot study in 2014 (Authors, 2015).

The surveys' main focus was on individual *target students*: the teachers reported on all students in their class (teacher ratings), the students described themselves (self-ratings) plus four randomly assigned classmates (peer ratings). The participating students acted both as raters and targets, the non-participating students were rated only by their teachers. On both occasions, the same raters and targets were paired together with the aid of personalised questionnaires: the raters' and the targets' names were printed on a tear-off strip. Several measures were taken to

guarantee the respondents anonymity. Supervising members of the project team reported that the students had been in a good mood after the survey (Authors, 2018).

5.2 Instruments

The questionnaire consisted of a general part (e.g. regarding teaching styles) and of a specific part that focused on the target students. This paper addresses the target-specific instruments exclusively; they were identical for teacher, peer and self-ratings except for the item wording (see Appendix A).

5.2.1 Deviant student behaviour

In the first part of the survey, we asked the respondents how frequently the target students had behaved deviantly in the preceding two weeks. The instrument comprised 18 items. The answering format consisted of six categories (“Never” to “5 times”) plus an option for free answers (“More frequently, namely: ...”). The factorial structure covered two dimensions (Authors, in preparation):

- *Undisciplined behaviour*: peer ratings: $\alpha = .79$, teacher ratings: $\alpha = .85$, self-ratings: $\alpha = .64$; 8 items, e.g. “[Name of the target] talked to another child during the lesson although the students were supposed to be quiet.”
- *Dissocial behaviour*: peer ratings: $\alpha = .87$, teacher ratings: $\alpha = .86$, self-ratings: $\alpha = .81$; 10 items, e.g. “[Name of the target] insulted another child in class.”

5.2.2 *Subjective perception of disturbance*

One week later, the respondents assessed the intensity of disturbance the targets had recently caused according to their subjective perception. They rated nine statements on a four-point rating scale (“Strongly disagree” to “Strongly agree”). Eight of these nine items were included in the factor analyses, which led to two dimensions (Authors, in preparation):

- *Affective perception of disturbance*: peer ratings: $\alpha = .87$, teacher ratings: $\alpha = .79$, self-ratings: $\alpha = .70$; 4 items, e.g. “[Name of the target] ... annoyed me.”
- *Cognitive perception of disturbance*: peer ratings: $\alpha = .92$, teacher ratings: $\alpha = .93$, self-ratings: $\alpha = .79$; 4 items, e.g. “[Name of the target] ... distracted me from the lesson.”

5.3 *The CT-C(M-1) modelling technique*

Following Eid et al. (2008) and Carretero-Dios et al. (2011), a *two-level CT-C(M-1) model* was set up and calculated in Mplus 8.0 (Muthén & Muthén, 2017). The model comprises four target-specific traits (undisciplined and dissocial behaviour, affective and cognitive perception of disturbance) which have been measured by three different methods (teacher, peer and self-ratings). The most important principles of the modelling are explained in this section. Further explanations follow in the results section and in Appendix B.

Level 2 (L2) is the level of the target students ($nL2 = 1677$) where the teacher ratings, the self-ratings, and the aggregated peer ratings have been modelled. *Level 1* (L1), by contrast, is the level of unique peer ratings ($nL1 = 5811$). Each target student received 3.47 peer ratings on average. These unique peer ratings are nested within targets – they have been aggregated at L2 into error-free random intercepts (true scores).

Figure 2 provides a *sketch of the model*: Ovals symbolise latent factors, boxes represent manifest indicators (item parcels) with tiny grey arrows illustrating the measurement errors. The symbols are labelled with acronyms; in each case, the first letter fits the constructs' denotation (e.g. “u”: undisciplined behaviour). Blue refers to teacher ratings, labelled with the letter “t”. Yellow refers to self-ratings, labelled with the letter “s”, and red refers to peer ratings, labelled with “p”. The trait factors are white as they represent the shared perspective of all raters. The labels' numbers refer to the item parcels which are numbered consecutively (see Appendix A). Unidirectional arrows illustrate factor loadings (λ), double arrows represent covariances (ψ).

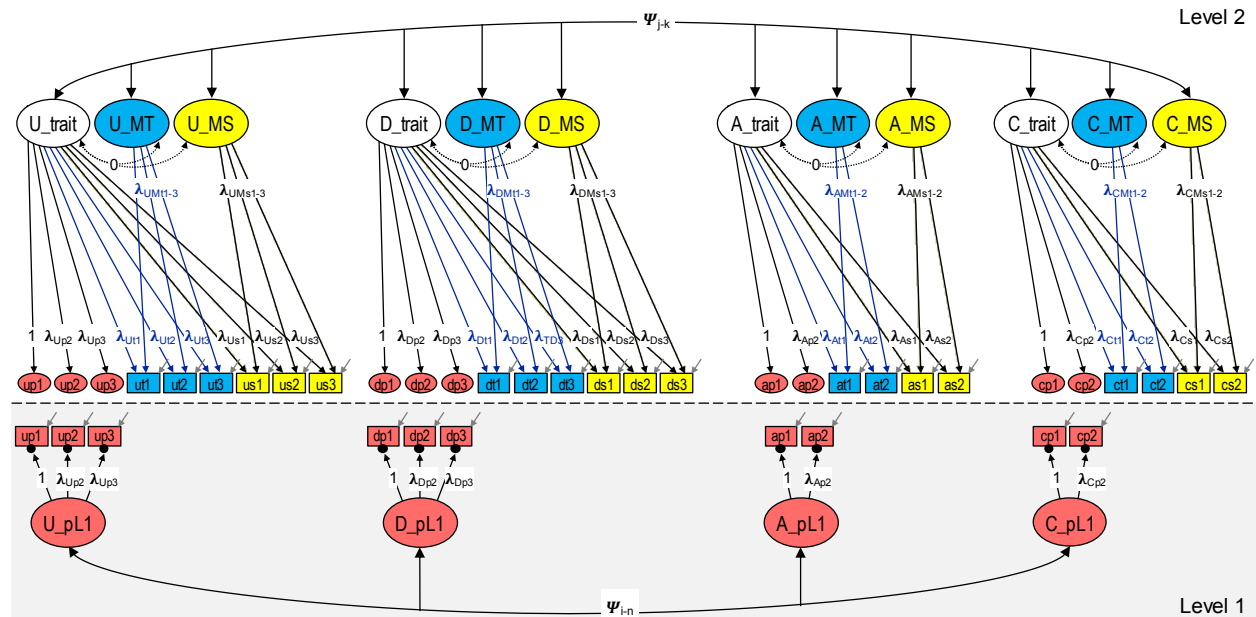


Figure 2. Sketch of the two-level CT-C(M-1) model.

I will explain the modelling in detail by an exemplary look at the construct “undisciplined behaviour”. The *white oval* “U_trait” at L2 symbolises the latent trait factor of the targets’ undisciplined behaviour. U_trait unites the proportion of variance that is shared across all types of rating and thus represents the raters’ common view on the targets’ behaviour. The *blue oval* “U_MT” stands for the method factor of the teacher ratings; the *yellow oval* “U_MS” represents

the method factor of the self-ratings. These method factors comprise the proportion of variance that is specific to those two types of rating (not shared with other types of rating). The *blue boxes* “*ut1–ut3*” symbolise the indicators of the teacher ratings, the *yellow boxes* “*us1–us3*” represent the indicators of the self-ratings. These indicators are influenced by the trait factor and by each according method factor. The influences are estimated by the factor loadings “ λ ”.

The *small red ovals* “*up1–up3*” at L2 illustrate the aggregated peer ratings. Unlike the teacher and self-ratings, these are not indicators at L2 but error-free random intercepts. No method factor was modelled for these aggregated peer ratings because they have been selected as the *reference method* with which the other types of rating are compared (comparison standard). It implies further that the peer ratings’ L2-scores are determined by the trait exclusively (no rater effect influences the aggregated peer ratings). The *red boxes* “*up1–up3*” at L1 represent the manifest indicators of the *unique* peer ratings. The black dots on the edge of these boxes indicate that they have been aggregated into random intercepts at L2. Each target student was rated by 3.47 peers on average. These unique peer ratings per target may differ from one another.

The *red oval* “*U_{pL1}*” at L1 illustrates a unique method factor of the unique peer ratings. This factor comprises the specific proportion of variance that is due to interindividual differences among the multiple peer raters per target (neither shared among peers, nor shared with other types of rating). It estimates the rater effect which influences the unique peer ratings.

Altogether, *U_{trait}* corresponds to the true scores per target, measured by the error-free aggregated peer ratings. In addition, *U_{trait}* encompasses the proportion of the teacher ratings and self-ratings that is consistent with the aggregated peer ratings. That is to say, *U_{trait}* is the *intersection of the different perspectives*. This intersubjectively consistent measurement can be

considered as an approximation to an objective information about the targets undisciplined behaviour. The method factors comprise the disagreement among the raters.

The *other three constructs* have been modelled in the same way. The white ovals illustrate the trait factors of dissocial behaviour (D_trait), affective perception of disturbance (A_trait), and cognitive perception of disturbance (C_trait). The blue ovals represent the teachers' method factors (_MT); the yellow ovals symbolise the self-ratings' method factors (_MS); the red ovals at L1 represent the peer ratings' unique method factors (_pL1). The red boxes at L1 stand for the unique peer ratings (dp1–dp3, ap1–ap2, cp1–cp2); the small red ovals at L2 illustrate the aggregated scores (labelled identically as the unique ratings). The blue boxes represent the teacher ratings (dt1–dt2, at1–at2, ct1–ct2); the yellow boxes represent the self-ratings (ds1–ds2, as1–as2, cs1–cs2).

Further measures have been adopted: The correlations between trait and method factors of the same trait-method unit are explicitly set to zero, because these factors are uncorrelated by definition (Eid et al., 2008). The factor loadings of the peer ratings are set equal at both levels to avoid cluster biases (Jak, Oort, & Dolan, 2013) which does not worsen the model fit (Chen, 2007). The 30 manifest indicators are item parcels. Parcelling was necessary for reducing the complexity of the model. This is justified because the analysis primarily aimed to estimate the relations between the constructs (T. D. Little, Cunningham, Shahar, & Widaman, 2002). Because the data are not normally distributed (Authors, 2017), the distinctive Mplus estimator with robust standard errors “MLR” (Finney & DiStefano, 2013) was applied. The MLR estimator uses all information available to model missing data (Muthén & Muthén, 2017).

6 Results

The two-level CT-C(M-1) model fits the data well ($\chi^2[\text{MLR}] = 1360.18$, $df = 372$, $p < .001$; RMSEA = .021; CFI = 0.93; SRMRL1 = 0.04; SRMRL2 = 0.04). Figure 3 shows the estimated standardised factor loadings and correlations. Because all included parameters are significant ($p < .05$), the usual marking with asterisks (*) has been omitted for the sake of clarity. Furthermore, only correlations with $|r| > 0.20$ are displayed (a complete list of all correlations is provided in Table 2).

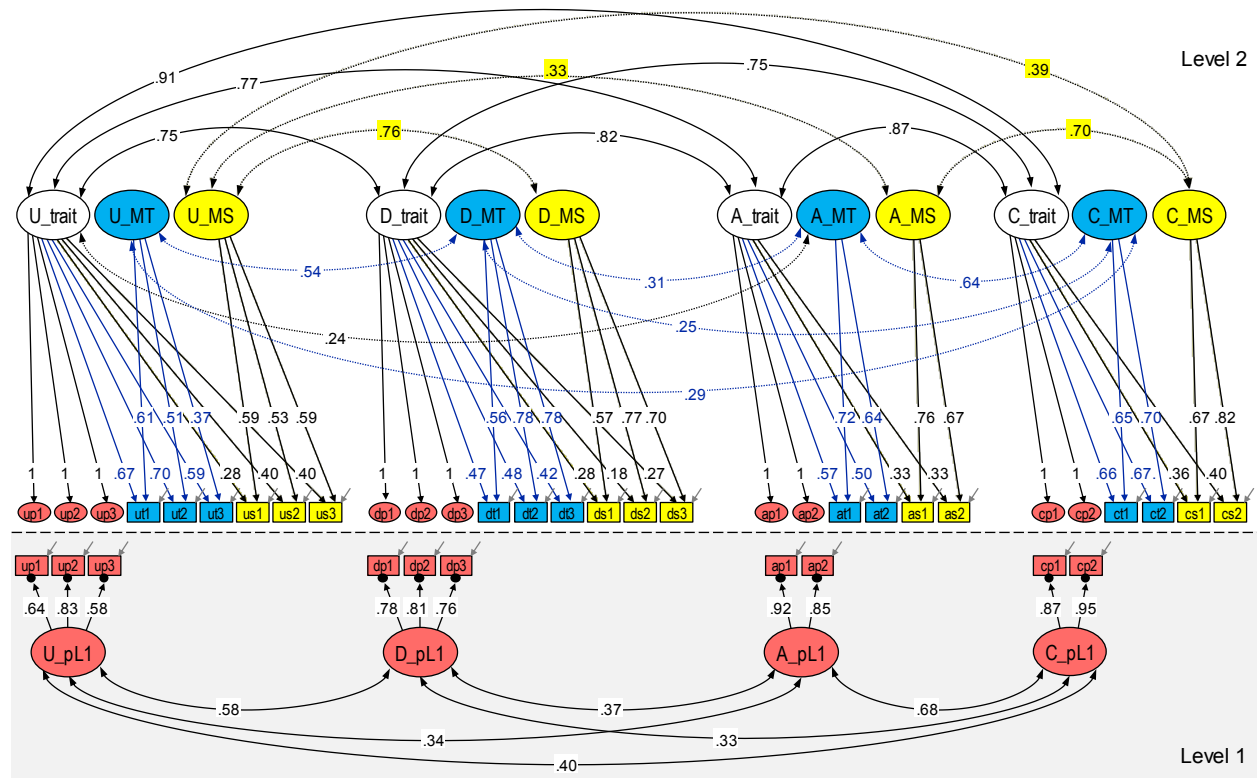


Figure 3. Standardised parameters according to the two-level CT-C(M-1) model estimation.

All factor loadings are significant and for the most part within an acceptable range. The standardised trait-factor loadings of the aggregated peer ratings amount to 1.00 because they are perfectly explained by the trait factor (measurement errors and rater effects are completely at L1). Compared to this, the freely estimated loadings of the teacher indicators on the trait factors

are lower ($.42 \leq \lambda \leq .70$). The loadings of the self-ratings are even lower ($.18 \leq \lambda \leq .40$). This is a first indication that the aggregated peer ratings (reference method) converge more strongly with the teacher ratings than with the self-ratings.

Table 1 presents further results. As for the *means (M)*, the target students on average only rarely showed undisciplined behaviour ($.68 \leq M \leq 1.68$) and even more rarely dissocial behaviour ($.22 \leq M \leq .44$). That the means are very low becomes evident if one keeps in mind that the rating scale was not limited to a fixed maximal number of incidents but included an option for free answers. Given these low mean incidences, it is not surprising that the target students on average were considered to be only little annoying ($.27 \leq M \leq .70$) and distracting ($.39 \leq M \leq .53$). The theoretical means of these two scales amount to 1.5.

The *unstandardized trait factor loadings* indicate level differences between the types of rating (Geiser, Eid, West, Lischetzke, & Nussbeck, 2012): loadings greater than 1.00 indicate that the pertaining self-ratings or teacher ratings are higher than the average peer rating (comparison standard); loadings lesser than 1.00 indicate lower self- or teacher ratings compared to the peer ratings. In sum, the teachers on average reported more incidents of undisciplined behaviours but fewer incidents of dissocial behaviours than the peers. Furthermore, the teachers on average reported a greater intensity of distraction but a lesser intensity of annoyance compared to the peers. The self-ratings on average are lower than the peer ratings as regards all four constructs.

Table 1. Factor Variance, Indicator-Specific Means, Unstandardized Factor Loadings, and Variance Components.

	Rater	Factor variance	Indicator	Trait factor loadings	Method factor loadings	Means univariate	Consistency observed	Specificity observed	Reliability
Undisciplined behaviour	Peers unique	.64	up1		1.00	.99	.21	.33	.54
			up2		1.17	.79	.30	.48	.78
			up3		1.10	.68	.17	.27	.45
	Peers aggregated	.40	up1	1.00					1.00
			up2	1.17					1.00
			up3	1.10					1.00
	Teacher	2.36	ut1	2.66	1.00	1.68	.45	.38	.83
			ut2	2.42	.72	1.24	.49	.26	.75
			ut3	3.14	.81	1.30	.35	.14	.49
	Self	.75	us1	.64	1.00	1.07	.08	.34	.42
			us2	.74	.72	.73	.16	.28	.44
			us3	.94	1.02	.68	.16	.34	.50
Dissozial behaviour	Peers unique	.51	dp1		1.00	.41	.14	.52	.66
			dp2		1.25	.44	.15	.56	.71
			dp3		.76	.34	.13	.49	.63
	Peers aggregated	.14	dp1	1.00					1.00
			dp2	1.25					1.00
			dp3	.76					1.00
	Teacher	.18	dt1	.96	1.00	.30	.22	.31	.53
			dt2	.86	1.23	.27	.23	.61	.84
			dt3	.70	1.15	.22	.17	.61	.78
	Self	.14	ds1	.49	1.00	.32	.08	.32	.40
			ds2	.66	2.81	.36	.03	.59	.62
			ds3	.44	1.13	.30	.07	.49	.56
Affective perception of disturbance	Peers unique	.43	ap1		1.00	.70	.23	.65	.87
			ap2		.91	.67	.20	.57	.77
	Peers aggregated	.15	ap1	1.00					1.00
			ap2	.91					1.00
	Teacher	.16	at1	.80	1.00	.35	.33	.52	.85
			at2	.63	.80	.27	.25	.40	.65
	Self	.15	as1	.43	1.00	.51	.11	.57	.68
			as2	.45	.93	.47	.11	.45	.56
Cognitive perception of disturbance	Peers unique	.34	cp1		1.00	.52	.20	.60	.80
			cp2		1.10	.50	.23	.70	.93
	Peers aggregated	.11	cp1	1.00					1.00
			cp2	1.10					1.00
	Teacher	.24	ct1	1.50	1.00	.53	.44	.42	.85
			ct2	1.54	1.10	.52	.44	.49	.93
	Self	.17	cs1	.66	1.00	.45	.13	.45	.58
			cs2	.68	1.15	.39	.16	.67	.83

Note. “Factor variance” indicated for teacher ratings, self-ratings and unique peer ratings relates to the (unique) method factors. As for the aggregated peer ratings (reference method), “factor variance” relates to the trait factor. The factor loadings are non-standardised values. The loadings of the unique peer ratings on the unique method factors are equivalent to the loadings of the aggregated peer ratings on the trait factors (no cluster bias). Reliability of the error-free aggregated peer ratings amounts to 1.00 because they are completely explained by the trait factor (the error is completely at L1).

Using the formulas proposed by Eid et al. (2008), indicator-specific variance components were manually calculated. *Consistency* (C) is a measure of interrater agreement that quantifies the intersection between the raters' distinct perspectives. It expresses the extent to which the aggregated peer ratings (reference method) explain the variance in the teacher ratings, in the self-ratings, and in the unique peer ratings. *Specificity* (S), by contrast, measures the rater effects: the coefficient is the proportion of variance in the ratings that is not consistent with the aggregated peer ratings. Technically speaking, it quantifies the overestimation or underestimation of other ratings compared to the average peer ratings. The *reliability* coefficient (ω) is the ratio of explained variance to total variance. This implies that unreliability ($\text{UnR} = 1 - \omega$) can be accounted for by the measurement error.

Consistency, specificity, and unreliability add up to 1.00 per indicator, and their proportions are to be interpreted in percentages when multiplied by 100. The components' meaning can be illustrated with an *example* from Table 1: the variance in the teacher ratings in *utl* is explained by the aggregated peer ratings to an extent of 45% (consistency). 38% of variance are method-specific, that is to say due to rater effects. The remaining 17% of variance can be accounted for by measurement errors (unreliability).

Some of the indicator-specific reliability coefficients are rather low. But as regards the *whole constructs*, reliability is considerably higher. Figures 4 to 7 illustrate the variance components (Varcomp) for all constructs per method. The bottom-most, grey part of the pillars represents unreliability of the constructs ($.06 \leq \text{UnR} \leq .29$). The green part above stands for consistency ($.07 \leq C \leq .54$). The uppermost, red part displays specificity ($.30 \leq S \leq .70$).

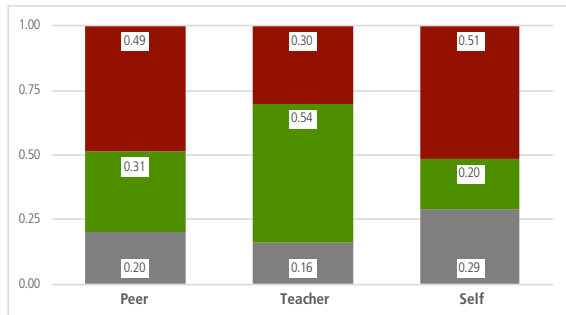


Figure 4. Varcomp of undisciplined behaviour.

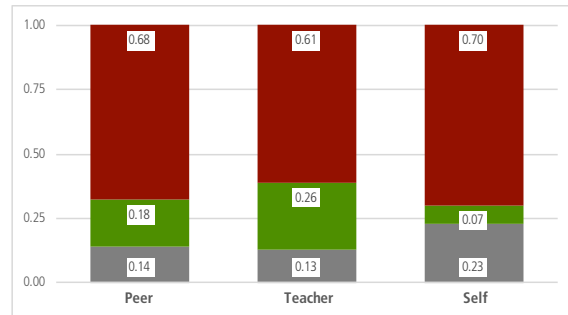


Figure 5. Varcomp of dissocial behaviour.

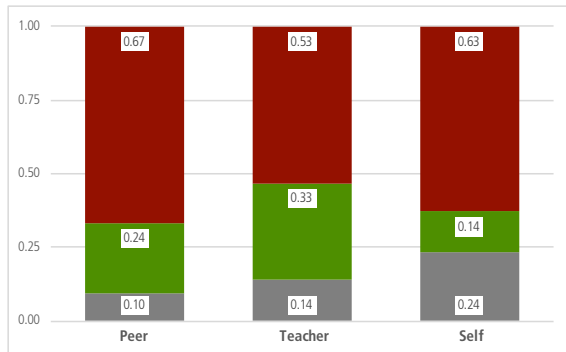


Figure 6. Varcomp of affective perception of disturbance.

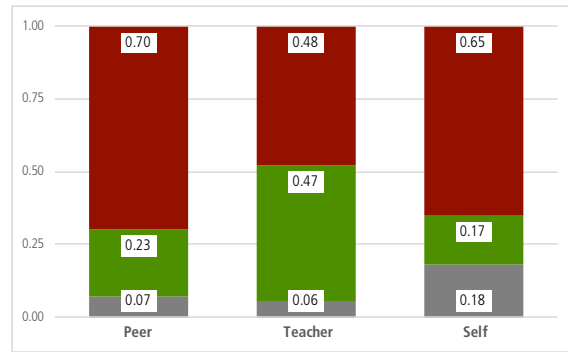
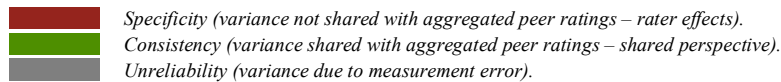


Figure 7. Varcomp of cognitive perception of disturbance.

Legend.



The *self-ratings* are marked by low consistency but high specificity over all constructs. This becomes most obvious in the construct “Dissocial behaviour”: only 7% of variance in the self-ratings can be explained by the aggregated peer ratings (reference method) whereas 70% of variance are rater-specific and can thus be explained by rater effects.

The variance in *teacher ratings* of the construct “Undisciplined behaviour” can be explained to an extent of 54% by the aggregated peer ratings (consistency). With respect to the construct “Dissocial behaviour”, by contrast, consistency only amounts to 26%. This discrepancy is unexpected because both constructs were measured with analogous low-inference scales that were assumed to lead to similar results. Moreover, we expected larger consistency coefficients for all deviance ratings. Also contrary to our expectations, the teacher ratings of the construct

“Cognitive perception of disturbance” are to an extent of 47% consistent with the aggregated peer ratings. Since this is a measure of subjective perceptions, we expected a lower consistency coefficient.

The *unique peer ratings* are to an extent of 18 to 31% consistent with the aggregated scores. In itself, the values in this range are rather low. What is even more surprising, however, is that all consistency coefficients of the unique peer ratings are lower than those of the teacher ratings. This means that the teacher ratings converge more strongly with the peers’ aggregated scores than the unique peer ratings – out of which the aggregated scores had originally been calculated.

Table 2 shows a *complete list of the correlations* between the latent factors. Significant results ($p < .05$) are marked by an asterisk (*). The table is made up of sectors (I to VII), which are described in the following.

Sector I shows the *correlations of the unique method factors*. All coefficients are positive and significant ($.33^* \leq r \leq .68^*$). Thus, the rater-specificity of the unique peer ratings is largely generalizable: if a peer rater overestimates a target in one construct (compared to the aggregated score), this peer tends to overestimate the same target in other constructs too. Or vice versa: underestimation in one construct coincides with underestimation in the other constructs. The correlations between the method factors of the teacher ratings (Sector IV) and between the method factors of the self-ratings (Sector VII) can be interpreted analogously.

Table 2. *Interfactor Correlations.*

		Level of unique peer ratings (L1)											
		U_pL1	D_pL1	A_pL1	C_pL1								
I													
Method factors of unique peer ratings	U_pL1	-											
	D_pL1	.58*	-										
	A_pL1	.34*	.37*	-									
	C_pL1	.40*	.33*	.68*	-								
		Target level (L2)											
		U_trait	D_trait	A_trait	C_trait								
II													
Trait factors	U_trait	-											
	D_trait	.75*	-										
	A_trait	.77*	.82*	-									
	C_trait	.91*	.75*	.87*	-								
		U_trait	D_trait	A_trait	C_trait	U_MT	D_MT	A_MT	C_MT				
III													
IV													
Method factors of teacher ratings	U_MT	-	-.13	-.19*	-.17	-							
	D_MT	.16*	-	.04	.12*	.54*	-						
	A_MT	.24*	.10	-	.16*	.19*	.31*	-					
	C_MT	.19*	.05	-.02	-	.29*	.25*	.64*	-				
		U_trait	D_trait	A_trait	C_trait	U_MT	D_MT	A_MT	C_MT	U_MS	D_MS	A_MS	C_MS
V													
VI													
VII													
Method factors of self-ratings	U_MS	-	-.06	-.12	-.05	-.01	-.02	.10	.07	-			
	D_MS	.05	-	-.01	-.01	-.06	-.02	-.01	-.11	.76*	-		
	A_MS	.12*	.08	-	.09*	.03	.02	.08	.08	.33*	.25	-	
	C_MS	.11*	.01	-.07*	-	.02	-.01	.06	.02	.39*	.21	.70*	-

Sector II includes the *correlations between the trait factors* which estimate the strength of the constructs' relation in the raters' common view (measured by the relatively small intersection of the different types of rating). All these correlations are positive and significant: the more frequently targets show undisciplined behaviour, the more frequently they behave in dissocial ways too ($r = .75^*$). Besides, the more frequently the targets show undisciplined behaviour, the more annoyed ($r = .77^*$) and distracted ($r = .91^*$) the respondents get. The same holds true with respect to dissocial behaviour ($.75^* \leq r \leq .82^*$). Although such findings were expected, the actual correlations are so high that they raise questions concerning discriminant validity.

Sector III displays the *correlations between the trait factors and the method factors* of the teacher ratings. The highest positive correlation indicates a tendency that targets who show increasingly frequent indiscipline according to the common view (U_trait) are perceived as increasingly ($r = .24^*$) annoying in their teacher's specific perspective which is not consistent with the aggregated peer ratings (A_MT). The correlations between the trait factors and the method factors of the self-ratings in Sector V can be interpreted in analogous ways, but they are rather low or not significant.

Sector IV lists the correlations between the method factors of the teacher ratings and the self-ratings. None of the coefficients is significant. This means that the teacher ratings and the self-ratings do not share any variance that they do not share with the aggregated peer ratings.

7 Discussion

7.1 Summary and conclusion

According to our theoretical framework, classroom disturbances consists of an objective core that not all persons involved perceive as equally disturbing (Authors, 2016a). We assume that deviant student behaviour commonly distracts and annoys teachers and students, yet the intensity of a perceived disturbance is affected by role-specific, individual, and contextual conditions. The X study investigates this interactionist phenomenon by means of a multi-perspective survey. Based on teacher, peer and self-ratings we measured the incidence of deviant student behaviour and the respondents' subjective perception of disturbance. The aim of this paper has been to analyse commonalities and differences between the three perspectives with a two-level CT-C(M-1) model. Four research questions have been pursued and can be answered as follows:

(Q-1) Consistency of the deviance ratings. Because of our low-inference instrument, we assumed to obtain rather unbiased ratings of deviant student behaviours. Therefore, we supposed that the teacher, peer and self-ratings would be largely consistent. In other words, we expected only minor rater effects. This hypothesis needs to be rejected. The extent to which the ratings are consistent merely amounts to 7 to 54%. As against this, rater effects explain between 30 and 70% of the measurements' variance. Only the teacher ratings of the construct "Undisciplined behaviour" are more consistent (54%) than specific (30%). In all other cases, the rater-specific divergences make up the larger proportion.

These unexpectedly large rater effects may in part be accounted for by the respondents' role-specific frames of perception (Scherzinger, Wettstein, & Wyler, 2017). As far as the self-ratings are concerned, the students may have underestimated the frequency of their own dissocial behaviour in terms of a self-serving strategy (70% specificity). This interpretation is supported by the low means of the self-ratings and by the unstandardized trait factor loadings of the self-ratings which are lower than those of the peer ratings.

The rater effects that influence the teacher ratings of dissocial behaviour (61% specificity) may be explained by differential opportunities in terms of its perceptibility: students probably hide dissocial behaviours due to fear of sanctions so that their teacher does not perceive every incident. This would explain why the teachers on average reported fewer incidents of dissocial behaviour compared to the average peer ratings according to the unstandardized trait factor loadings.

Undisciplined behaviour, by contrast, is socially less disapproved and therefore less hidden and better perceptible for the teachers (30% specificity). According to the unstandardized trait factor loadings, the teachers on average reported more incidents of indiscipline than the

peers. This may be due to the teachers' sensitivity to disciplinary issues in class. Maybe this implies further that the peers underestimated the indiscipline of their classmates.

(Q-2) Specificity of the subjective perception of disturbance. The assumption was that specificity is larger than consistency. This hypothesis is supported by the results: up to 70% of variance can be accounted for by rater effects. Solely regarding the teachers' cognitive perception of disturbance, the ratio is less pronounced than expected (47% consistency vs. 48% specificity). This implies that teachers and peer raters on average agreed to 47% on the intensity of distraction the targets had caused. This finding is surprising because it rests on highly subjective measurements so that the proportion of specificity was expected to be larger. In all other cases, specificity clearly outweighs consistency as expected. This becomes particularly evident in the self-ratings (63 to 65% specificity). The low means and the unstandardized trait factor loadings indicate that the students probably underestimated themselves in this respect compared to the average peer ratings, which might be once more a consequence of self-serving biases.

The rater effects that influence the teachers' affective perception of disturbance (53% specificity) can be outlined by the level difference revealed by the unstandardized trait factor loadings: the teachers on average described the targets as less annoying than the peers. The low means imply that the teachers described most students as not annoying. Maybe this is a consequence of their professional ethos.

(Q-3) Relation of deviance and perception of disturbance in the raters' common view. We expected medium to strong correlations between the four trait factors. This hypothesis is supported by the results: the more frequently the targets show deviant behaviour (according to the common view), the more annoyed and distracted the respondents get. Some of these correlations are so high, however, that they raise the question of whether the constructs

discriminate sufficiently. It has to be considered, for instance, whether the trait factors “Undisciplined behaviour” and “Cognitive perception of disturbance” should be merged into one super factor. This merging would mean to equate the incidence of indiscipline and the intensity of distraction as two undistinguishable aspects of the same phenomenon. We refrained from doing so because the merging of factors should result in theoretically explicable constructs (Kleinke, Schlüter, & Christ, 2017). In our view, the potential super factor would represent an obscure mixture of two distinct constructs. Furthermore, the constructs discriminate well in the raters’ specific perspectives (the method factors correlate much less strongly).

(Q-4) Generalisability of rater-effects. The assumption was that the rater effects tend to be similar across the different constructs. The results support this hypothesis. The analyses showed medium to high correlations between the rater-specific method factors ($.19^* \leq r \leq .76^*$). Raters who overestimated a target in one construct tended to overestimate this target regarding the other constructs as well (compared to the aggregated peer ratings). The same holds true vice versa: underestimation in one construct coincides with underestimation in the other constructs. This indicates that the rater effects are largely generalisable.

7.2 Limitations

The aggregated peer ratings have been selected as *reference method* because multiple ratings per target were available so that occasional exaggerations and trivialisations were assumed to even out. Therefore, the aggregated peer ratings were supposed to be a more precise method than the teacher ratings and self-ratings. On the basis of the X data, this assumption cannot be verified. A possible validation strategy in future studies could consist in complementing teacher and student ratings with observation data (Wettstein, Scherzinger, et al.,

2018). An alternative strategy might involve an attempt to increase the generalisability of the ratings through multiple data collections over a longer time span, for example by means of the experience sampling method (Zurbriggen, Venetz, & Hinni, 2018) and/or the direct behaviour rating approach (Christ, Riley-Tillman, & Chafouleas, 2009).

7.3 *Implications*

The analyses have revealed that rater effects clearly dominated the teachers' and the students' ratings of classroom disturbances. As far as *theory* is concerned, these findings confirm that the subjective perception of disturbance is susceptible to various influences. As a consequence, this emphasises known *methodological issues*: researchers need to consider which perspective (e.g. teacher, student, external observer) is the most appropriate for their aims (Kunter & Baumert, 2006). As an alternative, researchers might analyse commonalities and differences between distinct perspectives, e.g. with the CT-C(M-1) approach. Moreover, low-inference instruments can reduce rater effects to some extent. The consistency coefficients of our deviance scale amount up to 54% which is larger than usual interrater agreement in classroom research (Wagner et al., 2016).

Nevertheless, the large extent of the rater effects implies that the labelling of single “problem students” (Hunt et al., 1989) is questionable from an *ethical* point of view. As a final point, the *praxis* may benefit from such findings: teachers who are aware that classroom disturbances are an interactionist problem may reflect consciously on various intervention strategies (Thommen & Wettstein, 2007).

7.4 *Research perspectives*

The two-level CT-C(M-1) model that has been presented in this paper forms the basis for further analyses in the context of the X study. As a next step, we plan to investigate causes and preventions of classroom disturbances according to our theoretical model. However, these analyses need to overcome further methodological challenges (Koch, Holtmann, Bohn, & Eid, 2017).

The CT-C(M-1) modelling technique could henceforth serve as an advantageous tool in educational research, because it estimates the common view of teachers and students as well as rater-specific divergences. It could be applied, for instance, in studies on instructional quality (Pham et al., 2012) in order to investigate causes of interrater (dis-)agreement.

Keypoints

- The common view of teachers and students on classroom disturbances is only marginal.
- Teacher, peer and self-ratings of deviant student behaviour are consistent to an extent of 7 to 54%.
- Rater effects explain up to 70% of variance in the teachers' and the students' perception of disturbance.
- The rater effects are generalisable: teachers and students who overestimate a student in one aspect overestimate this student in other aspects too.

Acknowledgments

I would like to thank the X Foundation for supporting the X study (Project No. X) and the project leader X for his substantial support and his invaluable encouragement.

Furthermore, I would like to thank two anonymous reviewers for providing judicious comments that helped to improve this articles' quality.

References

- Author (2018)
 Author (in preparation)
 Authors (2016a)
 Authors (2016b)
 Authors (2017)
 Authors (in preparation)
 Authors (2015)
- Arbuckle, C., & Little, E. (2004). Teachers' perceptions and management of disruptive classroom behaviour during the middle years. *Australian Journal of Educational and Developmental Psychology*, 4, 59–70.
- Beaman, R., Wheldall, K., & Kemp, C. (2007). Recent research on troublesome classroom behaviour: A review. *Australasian Journal of Special Education*, 31(1), 45–60.
- Carretero-Dios, H., Eid, M., & Ruch, W. (2011). Analyzing multitrait-multimethod data with multilevel confirmatory factor analysis: An application to the validation of the State-Trait Cheerfulness Inventory. *Journal of Research in Personality*, 45(2), 153–164.
- Carroll, A., Houghton, S., Taylor, M., West, J., & List-Kerz, M. (2006). Responses to interpersonal and physically provoking situations. *Educational Psychology*, 26(4), 483–498.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504.
- Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Foundation for the development and use of Direct Behavior Rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, 34(4), 201–213.
- Crawshaw, M. (2015). Secondary school teachers' perceptions of student misbehaviour: A review of international research, 1983 to 2013. *Australian Journal of Education*, 59(3), 293–311.
- Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A., & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, 106(2), 569–583.

- Doyle, W. (2006). Ecological approaches to classroom management. In C. Evertson & C. Weinstein (Eds.), *Handbook of classroom management: Research, practice and contemporary issues* (pp. 97–125). Mahwah, NJ: Erlbaum.
- Eid, M., Lischetzke, T., Nussbeck, F., & Trierweiler, L. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, 8(1), 38–60.
- Eid, M., Nussbeck, F., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, 13(3), 230–253.
- Emmer, E. T., & Sabornie, E. J. (2015). *Handbook of classroom management* (2nd ed.). New York, NY: Routledge.
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation models. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439–492). Charlotte, NC: Information Age Publishing.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.
- Geiser, C., Eid, M., West, S. G., Lischetzke, T., & Nussbeck, F. W. (2012). A comparison of method effects in two confirmatory factor models for structurally different methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(3), 409–436.
- Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S., & Fisher, A. V. (2016). Off-task behavior in elementary school children. *Learning and Instruction*, 44, 128–143.
- Hamre, B. K., Pianta, R. C., Downer, J. T., & Mashburn, A. J. (2008). Teachers' perceptions of conflict with young students: Looking beyond problem behaviors. *Social Development*, 17(1), 115–136.
- Hempel-Jorgensen, A. (2009). The construction of the “ideal pupil” and pupils' perceptions of “misbehaviour” and discipline: Contrasting experiences from a low-socio-economic and a high-socio-economic primary school. *British Journal of Sociology of Education*, 30(4), 435–448.
- Hofer, M. (1986). *Sozialpsychologie erzieherischen Handelns [Social psychology of education]*. Goettingen: Hogrefe.
- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological methods*, 5(1), 64–86.
- Hoyt, W. T., & Kerns, M.-D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, 4(4), 403–424.
- Hunt, D., Carline, J., Tonesk, X., Yergan, J., Siever, M., & Loebel, J. (1989). Types of problem students encountered by clinical teachers on clerkships. *Medical Education*, 23(1), 14–18.
- Infantino, J., & Little, E. (2005). Students' perceptions of classroom behaviour problems and the effectiveness of different disciplinary methods. *Educational Psychology*, 25(5), 491–508.
- Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling*, 20(2), 265–282.
- Kleinke, K., Schlüter, E., & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus. Eine praktische Einführung [Structural equation modelling with Mplus. A practical introduction]* (2nd ed.). Berlin: de Gruyter.

- Koch, T., Holtmann, J., Bohn, J., & Eid, M. (2017). Explaining General and Specific Factors in Longitudinal, Multimethod, and Bifactor Models: Some Caveats and Recommendations. *Psychological Methods*. Advance online publication.
- Kokkinos, C. M. (2007). Job stressors, personality and burnout in primary school teachers. *British Journal of Educational Psychology*, 77, 229–243.
- Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9(3), 231–251.
- Little, E. (2005). Secondary school teachers' perceptions of students' problem behaviours. *Educational Psychology*, 25(4), 369–377.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9(2), 151–173.
- Makarova, E., Herzog, W., & Schönbächler, M.-T. (2014). Wahrnehmung und Interpretation von Unterrichtsstörungen aus Schülerperspektive sowie aus Sicht der Lehrpersonen [Perception and interpretation of classroom disturbances in the perspectives of students and teachers]. *Psychologie in Erziehung und Unterricht*, 61(2), 127–140.
- Montuoro, P., & Lewis, R. (2015). Student perceptions of misbehavior and classroom management. In E. T. Emmer & E. J. Sabornie (Eds.), *Handbook of classroom management* (2nd ed., pp. 344–362). New York, NY: Routledge.
- Müller, C. M., & Hofmann, V. (2016). Does being assigned to a low school track negatively affect psychological adjustment? A longitudinal study in the first year of secondary school. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 27(2), 95–115.
- Muthén, B. O., & Muthén, L. K. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Ohan, J., Visser, T. A. W., Strain, M. C., & Allen, L. (2011). Teachers' and education students' perceptions of and reactions to children with and without the diagnostic label "ADHD". *Journal of School Psychology*, 49, 81–105.
- Pham, G., Koch, T., Helmke, A., Schrader, F.-W., Helmke, T., & Eid, M. (2012). Do teachers know how their teaching is perceived by their pupils? *Procedia – Social and Behavioral Sciences*, 46, 3368–3374.
- Preuss-Lausitz, U. (2005). *Verhaltensauffällige Kinder integrieren. Zur Förderung der emotionalen und sozialen Entwicklung [Integrating children with behavioural problems. Fostering the emotional and social development]*. Weinheim: Beltz.
- Scherzinger, M., Wettstein, A., & Wyler, S. (2017). Unterrichtsstörungen aus der Sicht von Schülerinnen und Schülern und ihren Lehrpersonen. Ergebnisse einer Interviewstudie zum subjektiven Erleben von Störungen [Classroom disturbances in the perspectives of students and teachers. Findings of an interview study on the subjective experience of disturbances]. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 86(1), 70–83.
- Schönbächler, M.-T., Makarova, E., Herzog, W., Altin, Ö., Känel, S., Lehmann, V., & Milojevic, S. (2009). *Klassenmanagement und kulturelle Heterogenität: Ergebnisse 2. Forschungsbericht Nr. 37. [Classroom management and cultural heterogeneity: Results 2. Research report no. 37]*. Bern: Universität Bern.

- Sherman, J., Rasmussen, C., & Baydala, L. (2008). The impact of teacher factors on achievement and behavioural outcomes of children with Attention Deficit/Hyperactivity Disorder (ADHD): A review of the literature. *Educational Research*, 50(4), 347–360.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *The Journal of Educational Psychology*, 104(3), 743–762.
- Thommen, B., & Wettstein, A. (2007). Toward a Multi-Level-Analysis of Classroom Disturbances. *European Journal of School Psychology*, 5(1), 65-82.
- Venet, M., Zurbriggen, C., & Schwab, S. (2017). What do teachers think about their students' inclusion? Consistency of self- and teacher reports. Paper presented at the EARLI Biennial Main Conference. Education in the Crossroads of Economy and Politics, Tampere (FIN).
- Wagner, W., Gollner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, 108(5), 705–721.
- Wettstein, A. (2012). A conceptual frame model for the analysis of aggression in social interactions. *Journal of Social, Evolutionary, and Cultural Psychology*, 6(2), 141–157.
- Wettstein, A., Ramseier, E., & Scherzinger, M. (2018). Eine Mehrebenenanalyse zur Schülerwahrnehmung von Störungen im Unterricht der Klassen- und einer Fachlehrperson [A multilevel analysis of the students' perception of classroom disturbances during the instruction of class teachers and subject teachers]. *Psychologie in Erziehung und Unterricht*, 65(1), 1–16.
- Wettstein, A., Ramseier, E., Scherzinger, M., & Gasser, L. (2016). Unterrichtsstörungen aus Lehrer- und Schülersicht. Aggressive und nicht aggressive Störungen im Unterricht aus der Sicht der Klassen-, einer Fachlehrperson und der Schülerinnen und Schüler [Classroom disturbances in the perspectives of teachers and students. Aggressive and non-aggressive disturbances during instruction in the view of class teachers, subject teachers, and students]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48(4), 171–183.
- Wettstein, A., Scherzinger, M., & Ramseier, E. (2018). Unterrichtsstörungen, Beziehung und Klassenführung aus Lehrer-, Schüler- und Beobachterperspektive [Classroom disturbances, relationship, and classroom management in the perspectives of teachers, students and external observers]. *Psychologie in Erziehung und Unterricht*, 65(1), 58-74.
- Wolfe, E. W. (2004). Identifying rater effects using latent trait models. *Psychology Science*, 46, 35-51.
- Zevenbergen, R. (2001). Mathematics, social class, and linguistic capital. In B. Atweh, H. Forgasz, & B. Nebres (Eds.), *Sociocultural research on mathematics education* (pp. 201–215). Mahwah, NJ: Erlbaum.
- Zurbriggen, C., Venet, M., & Hinni, C. (2018). The quality of experience of students with and without special educational needs in everyday life and when relating to peers. *European Journal of Special Needs Education*, 33(2), 205–220.

Appendix A1. Undisciplined Behaviour: Item Wording and Parcelling

Item	Wording Peer Rating	Wording Teacher Rating	Wording Self-Rating
	Parcel up1	Parcel ut1	Parcel us1
u03	Talked to another child during the lesson although the students were supposed to be quiet.	Talked to another child during the lesson although the students were supposed to be quiet.	I talked to another child during the lesson although we were supposed to be quiet.
u06	Did not immediately obey when the teacher said something.	Did not immediately obey one of my instructions.	I did not immediately obey when the teacher said something.
u07	Talked angrily to the teacher.	Talked angrily to me.	I talked angrily to the teacher.
	Parcel up2	Parcel ut2	Parcel us2
u01	Did not properly participate in the lesson but did something else.	Did not properly participate in the lesson but did something else.	I did not properly participate in the lesson but did something else.
u04	Made noise during the lesson.	Made noise during the lesson.	I made noise during the lesson.
u08	Gave the teacher cheeky answers.	Gave me cheeky answers.	I gave the teacher cheeky answers.
	Parcel up3	Parcel ut3	Parcel us3
u02	Said something during the lesson without putting his/her hand up although we were required to do so.	Said something during the lesson without putting his/her hand up although it was required.	I said something during the lesson without putting my hand up although we were required to do so.
u05	Interrupted the teacher.	Interrupted me.	I interrupted the teacher.

Appendix A2. Dissocial Behaviour: Item Wording and Parcelling

Item	Wording Peer Rating	Wording Teacher Rating	Wording Self-Rating
	Parcel dp1	Parcel dt1	Parcel ds1
d12	Threw another child's things to the floor during the lesson.	Threw another child's things to the floor during the lesson.	I threw another child's things to the floor during the lesson.
d16	Was mean to another child (e.g. rumour spreading, bullying).	Was mean to another child (e.g. rumour spreading, bullying).	I was mean to another child (e.g. rumour spreading, bullying).
d17	Argued with another child (e.g. insulting, yelling).	Argued with another child (e.g. insulting, yelling).	I argued with another child (e.g. insulting, yelling).
d18	Hit or kicked another child.	Hit or kicked another child.	I hit or kicked another child.
	Parcel dp2	Parcel dt2	Parcel ds2
u09	Laughed at another child during the lesson.	Laughed at another child during the lesson.	I laughed at another child during the lesson.
d11	Insulted another child during the lesson.	Insulted another child during the lesson.	I insulted another child during the lesson.
d15	Hit or kicked another child during the lesson.	Hit or kicked another child during the lesson.	I hit or kicked another child during the lesson.
	Parcel dp3	Parcel dt3	Parcel ds3
d10	Talked nastily about another child during the lesson.	Talked nastily about another child during the lesson.	I talked nastily about another child during the lesson.
d13	Yelled at another child during the lesson.	Yelled at another child during the lesson.	I yelled at another child during the lesson.
d14	Pushed another child during the lesson.	Pushed another child during the lesson.	I pushed another child during the lesson.

Note: Items d16, d17, and d18 included an extra prompt: “How was it shortly before or after the lesson, for example on your way to school or during the break?” The intention behind this addition was to address dissocial forms of behaviour that had occurred outside the classroom but might have had consequences during the lesson, for example if the students involved had still been emotionally upset.

Appendix A3. Affective Perception of Disturbance: Item Wording and Parcelling

Item	Wording Peer Rating	Wording Teacher Rating	Wording Self-Rating
	Parcel ap1	Parcel at1	Parcel as1
a01r	... was always kind to me.	... was always kind to me. (r)	I was always kind to them. (r)
a04	... got on my nerves.	... got on my nerves.	I got on their nerves.
	Parcel ap2	Parcel at2	Parcel as2
a05	... annoyed me.	... annoyed me.	I annoyed them.
a09r	... was always friendly to me. (r)	... was always friendly to me. (r)	I was always friendly to them. (r)

Note. Items a01r and a09r had been positively worded so that the raters could describe the targets in a favourable way in order to prevent a negative stigmatisation of the target student due to the survey. The original answers were afterwards recoded into inverted values (r).

Appendix A4. Cognitive Perception of Disturbance: Item Wording and Parcelling

Item	Wording Peer Rating	Wording Teacher Rating	Wording Self-Rating
	Parcel cp1	Parcel ct1	Parcel cs1
p02	... kept me from learning during the lesson.	... kept me from teaching as such.	I disrupted the lesson.
p07	... distracted me from the lesson.	... distracted me from the lesson.	I distracted them from the lesson.
	Parcel cp2	Parcel ct2	Parcel cs2
p03	... disturbed me during the lesson.	... disturbed me during the lesson.	I disturbed the lesson.
p06	... disturbed my concentration.	... disturbed my concentration.	I disturbed their concentration.

Appendix B. Mplus Input File of the Two-level CT-C(M-1) Model

Mplus Syntax	Explanations
DATA: FILE = 0-UDAC-pts.dat;	
VARIABLE: NAMES = pidz up1 up2 up3 dp1 dp2 dp3 ap1 ap2 cp1 cp2 ut1 ut2 ut3 dt1 dt2 dt3 at1 at2 ct1 ct2 us1 us2 us3 ds1 ds2 ds3 as1 as2 cs1 cs2; MISSING = all (-9); USEVARIABLES = up1 up2 up3 dp1 dp2 dp3 ap1 ap2 cp1 cp2 ut1 ut2 ut3 dt1 dt2 dt3 at1 at2 ct1 ct2 us1 us2 us3 ds1 ds2 ds3 as1 as2 cs1 cs2;	
CLUSTER = pidz;	!unique peer ratings (L1) are clustered within targets (L2)
BETWEEN = ut1 ut2 ut3 dt1 dt2 dt3 at1 at2 ct1 ct2 us1 us2 us3 ds1 ds2 ds3 as1 as2 cs1 cs2;	!teacher and self-ratings are measured at target level (L2) !as they are mentioned in the BETWEEN statement, they will not be considered in the estimations at L1
	!the peer ratings are measured at L1 !as they are not mentioned in the BETWEEN statement, Mplus estimates them as random intercepts at L2
ANALYSIS: TYPE = TWOLEVEL; ESTIMATOR = MLR;	

...continues on next page...

Mplus Syntax	Explanations
MODEL:	
%WITHIN%	!this is the level of unique peer ratings (L1)
U_pL1 by up1(a) up2(b) up3(c);	!undisciplined behaviour – unique method factor of the peer ratings
D_pL1 by dp1(d) dp2(e) dp3(f);	!dissocial behaviour – unique method factor peer ratings
A_pL1 by ap1(g) ap2(h);	!affective perception of disturbance – unique method factor peer ratings
C_pL1 by cp1(i) cp2(k);	!cognitive perception of disturbance – unique method factor peer ratings
	!factor loadings of peer ratings are labelled (a) to (k) at L1 and at L2 to set them equal at both levels (no cluster bias)
%BETWEEN%	
up1 - cp2@0;	!the error of the peer ratings at L2 is set to zero because the error is completely at L1
U_trait by up1(a) up2(b) up3(c) ut1 ut2 ut3 us1 us2 us3;	!undisciplined behaviour – trait factor !peer ratings (reference method) !teacher ratings !self-ratings
D_trait by dp1(d) dp2(e) dp3(f) dt1 dt2 dt3 ds1 ds2 ds3;	!dissocial behaviour – trait factor !peer ratings (reference method) !teacher ratings !self-ratings
A_trait by ap1(g) ap2(h) at1 at2 as1 as2;	!affective perception of disturbance – trait factor !peer ratings (reference method) !teacher ratings !self-ratings
C_trait by cp1(i) cp2(k) ct1 ct2 cs1 cs2;	!cognitive perception of disturbance – trait factor !peer ratings (reference method) !teacher ratings !self-ratings
U_MT by ut1 ut2 ut3;	!undisciplined behaviour – method factor teacher ratings
U_MS by us1 us2 us3;	!undisciplined behaviour – method factor self-ratings
D_MT by dt1 dt2 dt3;	!dissocial behaviour – method factor teacher ratings
D_MS by ds1 ds2 ds3;	!dissocial behaviour – method factor self-ratings
A_MT by at1 at2;	!affective perception of disturbance – method factor teacher ratings
A_MS by as1 as2;	!affective perception of disturbance – method factor self-ratings
C_MT by ct1 ct2;	!cognitive perception of disturbance – method factor teacher ratings
C_MS by cs1 cs2;	!cognitive perception of disturbance – method factor self-ratings
U_trait with U_MT@0 U_MS@0; D_trait with D_MT@0 D_MS@0; A_trait with A_MT@0 A_MS@0; C_trait with C_MT@0 C_MS@0;	!trait factors are not allowed to correlate with method factors from the same trait-method unit
OUTPUT:	
SAMPSTAT STDYX TECH1	
TECH4;	

6.2. Curriculum Vitae

Boris Eckstein, geboren am 3. September 1982 in Zürich.

Derzeit wohnhaft in Winterthur, verheiratet, Vater zweier Töchter (*2009, *2011).

Ausbildung

- | | |
|-------------|---|
| 2011 – 2018 | <i>Doktoratsstudium</i> an der Universität Zürich im Fach Erziehungswissenschaft |
| 2003 – 2010 | <i>Studium</i> an der Philosophischen Fakultät der Universität Zürich
Hauptfach: Pädagogik
1. Nebenfach: Neuere deutsche Literaturwissenschaft
2. Nebenfach: Psychopathologie des Kindes- und Jugendalters |
| 1989 – 2002 | Volksschule in Uster und Gymnasium in Zürich |

Berufserfahrung

- | | |
|-------------|--|
| 2010 – 2018 | <i>Wissenschaftlicher Assistent, später wissenschaftlicher Mitarbeiter</i> von Prof. em. Dr. Kurt Reusser am Institut für Erziehungswissenschaft

<i>Dozent</i> im BA- und MA-Erziehungswissenschaft, Universität Zürich |
| 2003 – 2010 | <i>Sozialpädagoge</i> in verschiedenen Institutionen im Kanton Zürich (u.a. Kinderheime, Jugendpsychiatrie) |

Forschungserfahrung

- | | |
|-------------|---|
| 2014 – 2018 | <i>SUGUS</i>
Mitarbeit bei der Antragstellung an den SNF (Gesuchsteller: Prof. Reusser, bewilligter Beitrag: 405'040 CHF)
Verantwortung für die quantitative Teilstudie |
| 2010 – 2013 | <i>Heterogenitäts-Projekt</i>
Mitarbeit bei der Online-Befragung
Koautor des wissenschaftlichen Berichts |

6.3. Selbstständigkeitserklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig ohne unerlaubte Beihilfe verfasst und noch an keiner anderen Fakultät eingereicht zu haben.



Zürich, 14. November 2018

Boris Eckstein